(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/069232 A2

(51) International Patent Classification[7]: G06F 19/00, C07K 16/18, 14/285, 16/30, 14/575, 14/545, 14/475, 14/535, 14/50, 14/525, C12N 9/24, 9/64, 9/54, C07K 14/59, 14/415, 14/47, 16/18, 16/28

(21) International Application Number: PCT/EP02/01688

(22) International Filing Date: 18 February 2002 (18.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

01103954.2	19 February 2001 (19.02.2001)	EP
01105777.5	8 March 2001 (08.03.2001)	EP
01106538.0	15 March 2001 (15.03.2001)	EP
01106536.4	15 March 2001 (15.03.2001)	EP
01107012.5	20 March 2001 (20.03.2001)	EP
01106899.6	20 March 2001 (20.03.2001)	EP
01107568.6	27 March 2001 (27.03.2001)	EP
01110220.9	25 April 2001 (25.04.2001)	EP
01113228.9	30 May 2001 (30.05.2001)	EP
01124965.3	19 October 2001 (19.10.2001)	EP
01126859.6	12 November 2001 (12.11.2001)	EP

(71) Applicant *(for all designated States except US)*: **MERCK PATENT GMBH** [DE/DE]; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CARR, Francis, J.** [GB/GB]; Birchlea, The Holdings, Balmedie, Aberdeenshire AB23 8XU (GB). **CARTER, Graham** [GB/GB]; Longhills Cottage, By Newmachar, Aberdeenshire AB21 7XB (GB). **JONES, Tim** [GB/GB]; 27 Brickrow, Babraham, Cambridge CB2 4AJ (GB). **WILLIAMS, Stephen** [GB/GB]; Drumgowan Cottage, Auchleven, Insch, Aberdeenshire AB52 6QD (GB). **HAMILTON, Anita** [GB/GB]; 130 Kings Gate, Aberdeen AB15 4EQ (GB).

(74) Common Representative: **MERCK PATENT GMBH**; Frankfurter Strasse 250, 64293 Darmstadt (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR IDENTIFICATION OF T-CELL EPITOPES AND USE FOR PREPARING MOLECULES WITH REEDUCED IMMUNOGENICITY

(57) Abstract: This invention relates to a novel approach for identification of T-cell epitopes, that give rise to an immune reaction in a living host. By means of this novel method biological compounds can be generated which have a no or at least a reduced immunogenicity when exposed to the immune system of a given species and compared with the relevant non-modified entity. Thus the invention relates also to novel biological molecules, especially proteins and antibodies, obtained by the method according to the invention.

WO 02/069232 A2

METHOD FOR IDENTIFICATION OF T-CELL EPITOPES AND USE FOR PREPARING MOLECULES WITH REDUCED IMMUNOGENICITY

FIELD OF INVENTION

The invention relates to a novel approach of identifying T-cell epitopes that give rise to an immune reaction in a living host comprising calculation of potential T-cell epitope values for MHC Class II molecule binding sites in a peptide by means of computer-aided methods. The invention furthermore relates to methods for preparing biological molecules, above all proteins and antibodies which elicit an immunogenic response when exposed to a host, preferably a human. By means of this method molecules can be prepared which have no or a reduced immunogenicity when exposed to the immune system of a given species and compared with the relevant non-modified entity by reduction or removal of potential T-cell epitopes within the sequence of said originally immunogenic molecules. Thus, the invention relates also to novel biological molecules obtained by the method according to the invention.

BACKGROUND OF THE INVENTION

Therapeutic use of a number of peptides, polypeptides and proteins is curtailed because of their immunogenicity in mammals, especially humans. For example, when murine antibodies are administered to patients who are not immunosuppressed, a majority of such patients exhibit an immune reaction to the introduced foreign material by making human anti-murine antibodies (HAMA) (e.g. Schroff, R. W. et al (1985) *Cancer Res.* 45: 879-885; Shawler, D.L. et al (1985) *J. Immunol.* 135: 1530-1535). There are two serious consequences. First, the patient's anti-murine antibody may bind and clear the therapeutic antibody or immunoconjugate before it has a chance to bind, for example to a tumor, and perform its therapeutic function. Second, the patient may develop an allergic sensitivity to the murine antibody and be at risk of anaphylactic shock upon any future exposure to murine immunoglobulin.

Several techniques have been employed to address the HAMA problem and thus enable the use in humans of therapeutic monoclonal antibodies (see, for example, WO-A-8909622, EP-A-0239400, EP-A-0438310, WO-A-9109967). These recombinant DNA approaches have generally reduced the mouse genetic information in the final antibody construct whilst increasing the human genetic information in the final construct. Notwithstanding, the resultant "humanized" antibodies have, in several cases, still elicited an immune response in patients

(Issacs J.D. (1990) *Sem. Immunol.* 2: 449, 456; Rebello, P.R. et al (1999) *Transplantation* 68: 1417-1420).

A common aspect of these methodologies has been the introduction into the therapeutic antibody, usually of rodent origin, of amino acid residues, even significant tracts of amino acid residue sequences, identical to those present in human antibody proteins. For antibodies, this process is possible owing to the relatively high degree of structural (and functional) conservatism among antibody molecules of different species. For potentially therapeutic peptides, polypeptides and proteins, however, where no structural homologue may exist in the host species (e.g., human) for the therapeutic protein, such processes are not applicable. Furthermore, these methods have assumed that the general introduction of a human amino acid residue sequence will render the re-modeled antibody non-immunogenic. It is known, however, that certain short peptide sequences ("T-cell epitopes") can be released during the degradation of peptides, polypeptides or proteins within cells and subsequently be presented by molecules of the major histocompatability complex (MHC) in order to trigger the activation of T-cells. For peptides presented by MHC Class II, such activation of T-cells can then give rise to an antibody response by direct stimulation of B-cells to produce such antibodies. Accordingly, it would be desirable to eliminate potential T-cell epitopes from a peptide, polypeptide or a protein. Even proteins of human origin and with the same amino acid sequences as occur within humans can still induce an immune response in humans. Notable examples include therapeutic use of granulocyte-macrophage colony stimulating factor (Wadhwa, M. et al (1999) Clin. Cancer Res. 5: 1353-1361) and interferon alpha 2 (Russo, D. et al (1996) *Bri. J. Haem.* 94: 300-305; Stein, R. et al (1988) *New Engl. J. Med.* 318: 1409-1413).

The elimination of T-cell epitopes from proteins has been previously disclosed (see, for example, WO 98/52976, WO 00/34317). The general methods disclosed in the prior art comprise the following steps:

(a) Determining the amino acid sequence of the polypeptide or part thereof

(b) Identifying one or more potential T-cell epitopes within the amino acid sequence of the protein by any method including determination of the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays.

(c) Designing new sequence variants with one or more amino acids within the identified potential T-cell epitopes modified in such a way to substantially reduce or eliminate the

activity of the T-cell epitope as determined by the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays. Such sequence variants are created in such a way to avoid creation of new potential T-cell epitopes by the sequence variations unless such new potential T-cell epitopes are, in turn, modified in such a way to substantially reduce or eliminate the activity of the T-cell epitope.

(d) Constructing such sequence variants by recombinant DNA techniques and testing said variants in order to identify one or more variants with desirable properties.

Other techniques exploiting soluble complexes of recombinant MHC molecules in combination with synthetic peptides and able to bind to T-cell clones from peripheral blood samples from human or experimental animal subjects have been used in the art [Kern, F. et al (1998) *Nature Medicine* 4:975-978; Kwok, W.W. et al (2001) *TRENDS in Immunology* 22: 583-588] and may also be exploited in an epitope identification strategy.

The potential T-cell epitopes are generally defined as any amino acid residue sequence with the ability to bind to MHC Class II molecules. Such potential T-cell epitopes can be measured to establish MHC binding. Implicit in the term "T-cell epitope" is an epitope which when bound to MHC molecules can be recognized by the T-cell receptor, and which can, at least in principle, cause the activation of these T-cells. It is, however, usually understood that certain peptides which are found to bind to MHC Class II molecules may be retained in a protein sequence because such peptides are tolerated by the immune within the organism into which the final protein is administered.

The invention is conceived to overcome the practical reality that soluble proteins introduced into an autologous host with therapeutic intent, can trigger an immune response resulting in development of host antibodies that bind to the soluble protein. One example amongst others is interferon alpha 2 to which a proportion of human patients make antibodies despite the fact that this protein is produced endogenously [Russo, D. et al (1996) *Brit. J. Haem.* 94: 300-305; Stein, R. et al (1988) *New Engl. J. Med.* 318: 1409-1413]

MHC Class II molecules are a group of highly polymorphic proteins which play a central role in helper T-cell selection and activation. The human leukocyte antigen group DR (HLA-DR) are the predominant isotype of this group of proteins and the major focus of the present

invention. However, isotypes HLA-DQ and HLA-DP perform similar functions, hence the present invention is equally applicable to these. MHC HLA-DR molecules are homo-dimers where each "half" is a hetero-dimer consisting of α and ß chains. Each hetero-dimer possesses a ligand binding domain which binds to peptides varying between 9 and 20 amino acids in length, although the binding groove can accommodate a maximum of 9 - 11 amino acids. The ligand binding domain is comprised of amino acids 1 to 85 of the α chain, and amino acids 1 to 94 of the ß chain. DQ molecules have recently been shown to have an homologous structure and the DP family proteins are also expected to be very similar. In humans approximately 70 different allotypes of the DR isotype are known, for DQ there are 30 different allotypes and for DP 47 different allotypes are known. Each individual bears two to four DR alleles, two DQ and two DP alleles. The structure of a number of DR molecules has been solved and such structures point to an open-ended peptide binding groove with a number of hydrophobic pockets which engage hydrophobic residues (pocket residues) of the peptide [Brown et al *Nature* (1993) 364: 33; Stern et al (1994) *Nature* 368: 215]. Polymorphism identifying the different allotypes of class II molecule contributes to a wide diversity of different binding surfaces for peptides within the peptide binding grove and at the population level ensures maximal flexibility with regard to the ability to recognize foreign proteins and mount an immune response to pathogenic organisms.

There is a considerable amount of polymorphism within the ligand binding domain with distinct "families" within different geographical populations and ethnic groups. This polymorphism affects the binding characteristics of the peptide binding domain, thus different "families" of DR molecules will have specificities for peptides with different sequence properties, although there may be some overlap. This specificity determines recognition of Th-cell epitopes (Class II T-cell response) which are ultimately responsible for driving the antibody response to B-cell epitopes present on the same protein from which the Th-cell epitope is derived. Thus, the immune response to a protein in an individual is heavily influenced by T-cell epitope recognition which is a function of the peptide binding specificity of that individual's HLA-DR allotype. Therefore, in order to identify T-cell epitopes within a protein or peptide in the context of a global population, it is desirable to consider the binding properties of as diverse a set of HLA-DR allotypes as possible, thus covering as high a percentage of the world population as possible.

A principal factor in the induction of an immune response is the presence within the protein of peptides that can stimulate the activity of T-cell via presentation on MHC class II molecules. In order to eliminate or reduce immunogenicity, it is thus desirable to identify and remove T-cell epitopes from the protein.

The unmodified biological molecules can be produced by recombinant technologies, which are per se well known in the art, using a number of different host cell types.

However, there is a continued need for analogues of said biological molecules with enhanced properties. Desired enhancements include alternative schemes and modalities for the expression and purification of the said therapeutic, but also and especially, improvements in the biological properties of the protein. There is a particular need for enhancement of the *in vivo* characteristics when administered to the human subject. In this regard, it is highly desired to provide the selected biological molecule with reduced or absent potential to induce an immune response in the human subject. Such proteins would expect to display an increased circulation time within the human subject and would be of particular benefit in chronic or recurring disease settings such as is the case for a number of indications for said biological molecule.

SUMMARY OF THE INVENTION

The present invention relates, therefore, to two general aspects:

(a) a convenient and effective computational method for the identification and calculation of T-cell epitopes for a globally diverse number of MHC Class II molecules and, based on this knowledge, for designing and constructing new sequence variants of biological molecules with improved properties, and

(b) novel biologically active molecules to be administered especially to humans and in particular for therapeutic use; said biological molecules are according to this invention immunogenicly modified polypeptides, proteins or immunoglobulins (antibodies) produced according to the method of the invention, whereby the modification results in a reduced propensity for the biological molecule to elicit an immune response upon administration to the human subject.

In particular the invention relates to the modification of several generally well-known proteins and antibodies with high therapeutic benefit from human or non-human origin obtained by the method according to the invention to result in proteins that are substantially non-immunogenic or less immunogenic than any non-modified counterpart when used *in vivo*. The molecules

modified according to the novel method of this invention would expect to display an increased circulation time within the human subject and would be of particular benefit in chronic or recurring disease settings such as is the case for a number of indications. The present invention provides for, as specific embodiments and in order to demonstrate the efficacy of the inventive method, modified forms of said molecules that are expected to display enhanced properties *in vivo*. These molecules with modified immunogenicity, i.e. having a decreased immunogenic potential, can be used in pharmaceutical compositions. Such modified molecules are herein termed "immunogenicly" modified.

A method for identifying T-cell epitopes partially by means of computational means can be utilized to calculate theoretical T-cell epitope values and thus identify potential MHC Class II molecule binding peptides within a protein; wherein the binding site comprises a sequence of amino acid sites within the protein. The identified peptides can thereafter be modified without substantially reducing, and possibly enhancing, the therapeutic value of the protein. This computational method comprises selecting a region of the protein having a known amino acid residue sequence, sequentially sampling overlapping amino acid residue segments (windows) of predetermined uniform size and constituted by at least three amino acid residues from the selected region, calculating MHC Class II molecule binding score for each sampled segment, and identifying at least one of the sampled segments suitable for modification, based on the calculated MHC Class II molecule binding score for that segment. The overall MHC Class II binding score for the peptide can then be changed without substantially reducing therapeutic value of the protein.

The MHC Class II molecule binding score for a selected amino acid residue segment in one aspect of this invention is calculated by summing assigned values for each hydrophobic amino acid residue side chain present in the sampled amino acid residue segment of the peptide. To generate a graphical overview, the value of that sum can then be assigned to a single amino acid residue at about the midpoint of the segment. This procedure is repeated for each of the overlapping segments (windows) in the peptide region or regions of interest. The assigned value for each aromatic side chain present is about one-half of the assigned value for each hydrophobic aliphatic side chain. The hydrophobic aliphatic side chains are those present in valine, leucine, isoleucine and methionine. The aromatic side chains are those present in phenylalanine, tyrosine and tryptophan. The preferred assigned value for an aromatic side

chain is about 1 and for a hydrophobic aliphatic side chain is about 2. Other values can be utilized, however.

Thus, in a first aspect, the invention provides for a computational-based method suitable for identifying one or more potential T-cell epitope peptides within the amino acid sequence of a biological molecule by steps including determination of the binding of said peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays, said method comprises the following steps:

(a) selecting a region of the peptide having a known amino acid residue sequence;

(b) sequentially sampling overlapping amino acid residue segments of predetermined uniform size and constituted by at least three amino acid residues from the selected region;

(c) calculating MHC Class II molecule binding score for each said sampled segment by summing assigned values for each hydrophobic amino acid residue side chain present in said sampled amino acid residue segment; and

(d) identifying at least one of said segments suitable for modification, based on the calculated MHC Class II molecule binding score for that segment, to change overall MHC Class II binding score for the peptide without substantially the reducing therapeutic utility of the peptide.

In a specific embodiment, the invention relates to a method, wherein step (c) is carried out by using a Böhm scoring function modified to include 12-6 van der Waal's ligand-protein energy repulsive term and ligand conformational energy term by

(1) providing a first data base of MHC Class II molecule models;

(2) providing a second data base of allowed peptide backbones for said MHC Class II molecule models;

(3) selecting a model from said first data base;

(4) selecting an allowed peptide backbone from said second data base;

(5) identifying amino acid residue side chains present in each sampled segment;

(6) determining the binding affinity value for all side chains present in each sampled segment; and optionally

(7) repeating steps (1) through (5) for each said model and each said backbone.

In a further embodiment the binding score for each sampled sequence is calculated by (i) providing a first data base of MHC Class II molecule models; (ii) providing a second data base of allowed peptide backbones for said MHC Class II molecule models; (iii) providing a third database of allowed amino acid side chain conformations for each of the twenty amino acids at each position of each backbone; (iv) selecting a model from said first data base; (v) selecting an allowed peptide backbone from said second data base; (vi) identifying amino acid residue side chains present in each sampled segment together with their allowed conformations from said third database; (vii) determining the optimum binding affinity value for all side chains present in each sampled segment in each allowed conformation; (viii) repeating steps (v) through (vii) for each said backbone and determining the optimum binding score; and (ix) repeating steps (iv) through (viii) for each said model.

It should be understood that the three databases described above can be combined into one database or any two databases can be combined to provide a combined database.

The length of the amino acid residue segments to be sampled can vary. Preferably, the sampled amino acid residue segments are constituted by about 10 to about 15 amino acid residues, more preferably about 13 amino acid residues.

The sampled amino acid residue segments can be overlapping to a varying degree. Preferably, the sampled amino acid residue segments overlap substantially. Most preferably, consecutive sampled amino acid residue segments overlap one another by all but one amino acid residue. That is, in an amino acid residue segment having n residues, n-1 residues are overlapped by the next consecutive sampled amino acid residue segment.

Thus, in more detail, the invention relates furthermore to the following further preferred embodiments:

- an accordingly specified method, wherein the assigned value for each aromatic side chain is about one-half of the assigned value for each hydrophobic aliphatic side chain;
- an accordingly specified method, wherein the sampled amino acid residue segment is constituted by 13 amino acid residues;
- an accordingly specified method, wherein consecutive sampled amino acid residue segments overlap by one to five amino acid residues;
- an accordingly specified method, wherein consecutive sampled amino acid residue segments overlap one another substantially;

- an accordingly specified method, wherein all but one of amino acid residues in consecutive sampled amino acid residue segments overlap.

In a second basic aspect, the present invention provides, modified forms of different biological molecules with one or more T-cell epitopes removed, wherein said modification may be achieved by the methods described above and in the claims. The molecules can also be produced by the methods as described in the above-cited prior art, however, the molecules obtained by the methods of this invention show enhanced properties. In the prior art teachings, predicted T-cell epitopes are removed by the use of judicious amino acid substitution within the primary sequence of the therapeutic antibody or non-antibody protein of both non-human and human derivation.

The present invention provides for modified forms of proteins and immunoglobulins that are expected to display enhanced properties *in vivo*.

Therefore, it is an object of the invention to provide a method for preparing an immunogenicly modified biological molecule derived from a parent molecule, wherein the modified molecule has an amino acid sequence different from that of said parent molecule and exhibits a reduced immunogenicity relative to the parent molecule when exposed to the immune system of a given species; said method comprises: (i) determining the amino acid sequence of the parent biological molecule or part thereof; (ii) identifying one or more potential T-cell epitopes within the amino acid sequence of the protein by any method including determination of the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays, (iii) designing new sequence variants by alteration of at least one amino acid residue within the originally identified T-cell epitope sequences, said variants are modified in such a way to substantially reduce or eliminate the activity or number of the T-cell epitope sequences and / or the number of MHC allotypes able to bind peptides derived from said biological molecule as determined by the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays or by binding of peptide-MHC complexes to T-cells, (iv) constructing such sequence variants by recombinant DNA techniques and testing said variants in order to identify one or more variants with desirable properties, and (v) optionally repeating steps (ii) – (iv), wherein the identification of T-cell epitope sequences according to step (ii) is achieved by a method as specified above and below.

Specific embodiments of step (iii) according to the invention relate to the following summarized steps:

- an accordingly specified method, wherein 1 – 9 amino acid residues in any of the originally present T-cell epitope sequences are altered;
- an accordingly specified method, wherein one amino acid residues in any of the originally present T-cell epitope sequences is altered;
- an accordingly specified method, wherein the amino acid alteration is made with reference to an homologous protein sequence and or to *in silico* modeling techniques.
- an accordingly specified method, wherein the alteration of the amino acid residues is substitution, deletion or addition of originally present amino acid(s) residue(s) by other amino acid residue(s) at specific position(s).
- an accordingly specified method, wherein additionally further alteration, preferably by substitution, addition or deletion of specific amino acid(s), is conducted to restore biological activity of said biological molecule.

With the exception of step (ii) the other steps of the method disclosed can be achieved by methods and techniques which are well known for skilled workers. Since the modified biological molecules are prepared preferably by recombinant technologies corresponding DNA constructs which were deduced from the amino acid sequence after having completed the exchange of amino acid residues identified by the method of step (i). The recombinant techniques used herein are well known in the art (e.g. Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, USA).

The biological molecule obtained according to the invention is preferably a peptide, a protein, an antibody, an antibody fragment, or a fusion protein. The invention includes furthermore modifications, variants, mutations, fragments, derivatives, non-, partially- or completely glycosylated forms of said molecules having the same or similar biological and / or pharmacological activity.

Although the method disclosed in this invention is not limited to specific biological molecules, it is a specific embodiment of the invention to provide preferably molecules which are known in the art and show a therapeutic benefit and value. Thus it is a further object of the invention to provide an immunogenicly modified biological molecule derived from a parent molecule, wherein the modified molecule has an amino acid sequence different from that of said parent

molecule and exhibits a reduced immunogenicity relative to the parent molecule when exposed to the immune system of a given species, obtained by a method according to the invention as disclosed in detail above and below.

The biological molecules of special interest obtained by said method are selected from the groups:

(a) monoclonal antibodies:

anti- 40kD glycoprotein antigen antibody KS 1/4 ,

anti- GD2 antibody 14.18

anti-Her2 antibody 4D5 (murine) and humanized version (Herceptin®),

anti-Her1 (EGFR) antibody c225 and h425

anti- IL-2R (anti-Tac) antibody (Zenapax®),

anti- CD52 antibody (CAMPATH®);

anti-CD20 antibodies (C2B8, Rituxan®; Bexxar®)

antibody directed to the human C5 complement protein

(b) human proteins:

sTNF-R1, sTNF-R2, sTNFR-Fc (Enbrel®),

protein C, acrp30, ricin A, CNTFR ligands

subtilisin, GM-CSF, human follicle stimulating hormone (h-fsh)

ß-glucocerebrosidase, GLP-1, apolipoprotein A1,

leptin (human obesity protein), KGF, G-CSF,

BDNF, EPO, Il-1R antagonist.

The third basic aspect of the present invention relates to the T-cell epitope sequences that derive from the parent immunogenicly non-modified biological molecules. These epitopes are preferably 13mer petides. Within these peptides sequences having 9 consecutive amino acid residues are preferred. Thus it is another object of the invention to provide access to such epitopes and sequences. In more detail the invention relates to:

- a use of a potential T-cell epitope peptide within the amino acid sequence of a parent immunogenicly non-modified biological molecule identified according to any of the methods as described for preparing a biological molecule with reduced immunogenicity having the same biological activity;
- a corresponding use of a potential T-cell epitope peptide, wherein said T-cell epitope is a 13mer peptide;

- a use of a peptide sequence consisting of at least 9 consecutive amino acid residues of a 13mer T-cell epitope as specified above for preparing a biological molecule with reduced immunogenicity having the same biological activity as compared with the parent non-modified molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a flow chart illustrating one aspect of the present computational method;

FIGURE 2 is a flow chart illustrating a database generation for a computational method embodying the present invention;

FIGURE 3 is a flow chart illustrating database interrogation for profiling a peptide for potential T-cell epitopes;

FIGURE 4 is a further flow chart illustrating the computational method.

FIGURE 5 is a plot of T-cell epitope likelihood index versus amino acid residue coordinates (positions) of glutamic acid decarboxylase (MW: 65000) isoform (GAD 65);

FIGURE 6 is a plot of T-cell epitope likelihood index versus amino acid residue coordinates (positions) for erythropoietin (EPO);

FIGURE 7 is a plot of T-cell epitope likelihood index versus amino acid residue coordinates (positions) for humanized anti-A33 monoclonal antibody light chain; and

FIGURE 8 is a plot of T-cell epitope likelihood index versus amino acid residue coordinates (positions) for humanized anti-A33 monoclonal antibody heavy chain.

In the foregoing FIGURES 5 –8, the solid line (—) depicts a T-cell epitope index calculated by a computational method in accordance with the flow chart shown in FIGURE 1, and the dotted line (•••••) depicts the predicted number of T-cell epitopes calculated in accordance with the computational method in accordance with the flow chart shown in FIGURE 3 according to another aspect of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The term "T-cell epitope" means according to the understanding of this invention an amino acid sequence which is able to bind with reasonable efficiency MHC class II molecules (or their equivalent in a non-human species), able to stimulate T-cells and / or also to bind (without necessarily measurably activating) T-cells in complex with MHC class II.

The term "peptide" as used herein and in the appended claims, is a compound that includes two or more amino acids. The amino acids are linked together by a peptide bond (defined

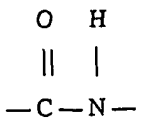
herein below). There are 20 different naturally occurring amino acids involved in the biological production of peptides, and any number of them may be linked in any order to form a peptide chain or ring. The naturally occurring amino acids employed in the biological production of peptides all have the L-configuration. Synthetic peptides can be prepared employing conventional synthetic methods, utilizing L-amino acids, D-amino acids, or various combinations of amino acids of the two different configurations. Some peptides contain only a few amino acid units. Short peptides, e.g., having less than ten amino acid units, are sometimes referred to as "oligopeptides". Other peptides contain a large number of amino acid residues, e.g. up to 100 or more, and are referred to as "polypeptides". By convention, a "polypeptide" may be considered as any peptide chain containing three or more amino acids, whereas a "oligopeptide" is usually considered as a particular type of "short" polypeptide. Thus, as used herein, it is understood that any reference to a "polypeptide" also includes an oligopeptide. Further, any reference to a "peptide" includes polypeptides, oligopeptides, and proteins. Each different arrangement of amino acids forms different polypeptides or proteins. The number of polypeptides–and hence the number of different proteins–that can be formed is practically unlimited.

The term "*less or reduced immunogenic(ity)*" used before and thereafter is a relative term and relates to the immunogenicity of the respective original source molecule when exposed in vivo to the same type of species compared with the molecule modified according to the invention. The term "modified protein" as used according to this invention describes a protein which has reduced number of T-cell epitopes and elicits therefore a reduced immunogenicity relative to the parent protein when exposed to the immune system of a given species. The term "non-modified protein" as used according to this invention describes the "parent" protein as compared to the "modified protein" and has a larger number of T- cell epitopes and, therefore, an enhanced immunogenicity relative to the modified protein when exposed to the immune system of a given species.

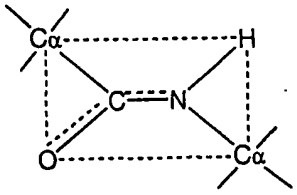
"Alpha carbon (Cα)" is the carbon atom of the carbon-hydrogen (CH) component that is in the peptide chain. A "side chain" is a pendant group to Cα that can comprise a simple or complex group or moiety, having physical dimensions that can vary significantly compared to the dimensions of the peptide.

T-cell epitopes can be identified by the computational method of the current invention by consideration of amino acid residues important for the binding of a particular T-cell epitope to

MHC Class II molecules. Once identified, potential T-cell epitopes can be removed or obliterated from an amino acid residue sequence by alteration, such as mutation, of key amino acid residues in that sequence. Any modification made to the sequence of a peptide in a region which is likely to contain T-cell epitopes, by deletion, addition or substitution, resulting in a relatively lower overall binding score will have the effect of rendering the amino acid residue sequence less immunogenic. In some instances, it may be desirable to enhance the binding of certain peptides to MHC Class II molecules. For example, it has been proposed that tolerance to certain autoantigens can be reinstated in individuals suffering from autoimmune disease if such individuals are treated with peptide analogues of regions of the autoantigen that are known to contain T-cell epitopes. The natural epitope usually has moderate affinity for MHC Class II molecules, whereas the peptide analogue is made such that it has a relatively higher affinity for MHC Class II molecules. This high affinity is important in either promoting immune surveillance to clear such T-cells presenting this high affinity epitope, or for them to become anergised. This modification to a T-cell epitope can also be made at the protein level of the peptide, and the entire protein administered as a therapeutic. There are a number of factors that play important roles in determining the total structure of a protein or polypeptide. First, the peptide bond, i.e., that bond which joins the amino acids in the chain together, is a covalent bond. This bond is planar in structure, essentially a substituted amide. An "amide" is any of a group of organic compounds containing the grouping:



The planar peptide bond linking Cα of adjacent amino acids may be represented as depicted below:



Because the O=C and the C-N atoms lie in a relatively rigid plane, free rotation does not occur about these axes. Hence, a plane schematically depicted by the interrupted line is sometimes

referred to as an "amide" or "peptide plane" plane wherein lie the oxygen (O), carbon (C), nitrogen (N), and hydrogen (H) atoms of the peptide backbone. At opposite corners of this amide plane are located the Cα atoms. Since there is substantially no rotation about the O=C and C-N atoms in the peptide or amide plane, a polypeptide chain thus comprises a series of planar peptide linkages joining the Cα atoms.

A second factor that plays an important role in defining the total structure or conformation of a polypeptide or protein is the angle of rotation of each amide plane about the common Cα linkage. The terms "angle of rotation" and "torsion angle" are hereinafter regarded as equivalent terms. Assuming that the O, C, N, and H atoms remain in the amide plane (which is usually a valid assumption, although there may be some slight deviations from planarity of these atoms for some conformations), these angles of rotation define the N and R polypeptide's backbone conformation, i.e., the structure as it exists between adjacent residues. These two angles are known as ϕ and ψ. A set of the angles ϕ_1, ψ_1, where the subscript i represents a particular residue of a polypeptide chain, thus effectively defines the polypeptide The conventions used in defining the ϕ, ψ angles, i.e., the reference points at which the amide planes form a zero degree angle, and the definition of which angle is ϕ, and which angle is ψ, for a given polypeptide, are defined in the literature. See, e.g,, Ramachandran et al. *Adv. Prot. Chem.* 23:283-437 (1968), at pages 285-94, which pages are incorporated herein by reference.

The present method can be applied to any protein, and is based in part upon the discovery that in humans the primary Pocket 1 anchor position of MHC Class II molecule binding grooves has a well designed specificity for particular amino acid side chains. The specificity of this pocket is determined by the identity of the amino acid at position 86 of the beta chain of the MHC Class II molecule. This site is located at the bottom of Pocket 1 and determines the size of the side chain that can be accommodated by this pocket (Marshall, K.W., (1994), *J. Immunol.*, 152:4946-4956). If this residue is a glycine, then all hydrophobic aliphatic and aromatic amino acids (hydrophobic aliphatics being: valine, leucine, isoleucine, methionine and aromatics being: phenylalanine, tyrosine and tryptophan) can be accommodated in the pocket, a preference being for the aromatic side chains. If this pocket residue is a valine, then the side chain of this amino acid protrudes into the pocket and restricts the size of peptide side chains that can be accommodated such that only hydrophobic aliphatic side chains can be accommodated. Therefore, in an amino acid residue sequence, wherever an amino acid with a hydrophobic aliphatic or aromatic side chain is found, there is the potential for a MHC Class II

restricted T-cell epitope to be present. If the side-chain is hydrophobic aliphatic, however, it is approximately twice as likely to be associated with a T-cell epitope than an aromatic side chain (assuming an approximately even distribution of Pocket 1 types throughout the global population).

A computational method embodying the present invention profiles the likelihood of peptide regions to contain T-cell epitopes as follows:

(1) The primary sequence of a peptide segment of predetermined length is scanned, and all hydrophobic aliphatic and aromatic side chains present are identified. (2)The hydrophobic aliphatic side chains are assigned a value greater than that for the aromatic side chains; preferably about twice the value assigned to the aromatic side chains, e.g., a value of 2 for a hydrophobic aliphatic side chain and a value of 1 for an aromatic side chain. (3) The values determined to be present are summed for each overlapping amino acid residue segment (window) of predetermined uniform length within the peptide, and the total value for a particular segment (window) is assigned to a single amino acid residue at an intermediate position of the segment (window), preferably to a residue at about the midpoint of the sampled segment (window). This procedure is repeated for each sampled overlapping amino acid residue segment (window). Thus, each amino acid residue of the peptide is assigned a value that relates to the likelihood of a T-cell epitope being present in that particular segment (window). (4) The values calculated and assigned as described in Step 3, above, can be plotted against the amino acid coordinates of the entire amino acid residue sequence being assessed. (5) All portions of the sequence which have a score of a predetermined value, e.g., a value of 1, are deemed likely to contain a T-cell epitope and can be modified, if desired.

This particular aspect of the present invention provides a general method by which the regions of peptides likely to contain T-cell epitopes can be described. Modifications to the peptide in these regions have the potential to modify the MHC Class II binding characteristics.

According to another aspect of the present invention, T-cell epitopes can be predicted with greater accuracy by the use of a more sophisticated computational method which takes into account the interactions of peptides with models of MHC Class II alleles.

The computational prediction of T-cell epitopes present within a peptide according to this particular aspect contemplates the construction of models of at least 42 MHC Class II alleles based upon the structures of all known MHC Class II molecules and a method for the use of these models in the computational identification of T-cell epitopes, the construction of libraries of peptide backbones for each model in order to allow for the known variability in

relative peptide backbone alpha carbon (Cα) positions, the construction of libraries of amino-acid side chain conformations for each backbone dock with each model for each of the 20 amino-acid alternatives at positions critical for the interaction between peptide and MHC Class II molecule, and the use of these libraries of backbones and side-chain conformations in conjunction with a scoring function to select the optimum backbone and side-chain conformation for a particular peptide docked with a particular MHC Class II molecule and the derivation of a binding score from this interaction.

Models of MHC Class II molecules can be derived via homology modeling from a number of similar structures found in the Brookhaven Protein Data Bank ("PDB"). These may be made by the use of semi-automatic homology modeling software (Modeller, Sali A. & Blundell TL., 1993. *J. Mol Biol* 234:779-815) which incorporates a simulated annealing function, in conjunction with the CHARMm force-field for energy minimisation (available from Molecular Simulations Inc., San Diego, Ca.). Alternative modeling methods can be utilized as well.

The present method differs significantly from other computational methods which use libraries of experimentally derived binding data of each amino-acid alternative at each position in the binding groove for a small set of MHC Class II molecules (Marshall, K.W., *et al.*, *Biomed. Pept. Proteins Nucleic Acids*, 1(3):157-162) (1995) or yet other computational methods which use similar experimental binding data in order to define the binding characteristics of particular types of binding pockets within the groove, again using a relatively small subset of MHC Class II molecules, and then 'mixing and matching' pocket types from this pocket library to artificially create further 'virtual' MHC Class II molecules (Sturniolo T., et al., *Nat. Biotech*, 17(6): 555-561 (1999). Both prior methods suffer the major disadvantage that, due to the complexity of the assays and the need to synthesize large numbers of peptide variants, only a small number of MHC Class II molecules can be experimentally scanned. Therefore the first prior method can only make predictions for a small number of MHC Class II molecules. The second prior method also makes the assumption that a pocket lined with similar amino-acids in one molecule will have the same binding characteristics when in the context of a different Class II allele and suffers further disadvantages in that only those MHC Class II molecules can be 'virtually' created which contain pockets contained within the pocket library. Using the modeling approach described herein, the structure of any number and type of MHC Class II molecules can be deduced, therefore alleles can be specifically selected to be representative of the global population. In addition, the number of MHC Class II molecules scanned can be

increased by making further models further than having to generate additional data via complex experimentation.

The use of a backbone library allows for variation in the positions of the Cα atoms of the various peptides being scanned when docked with particular MHC Class II molecules. This is again in contrast to the alternative prior computational methods described above which rely on the use of simplified peptide backbones for scanning amino-acid binding in particular pockets. These simplified backbones are not likely to be representative of backbone conformations found in 'real' peptides leading to inaccuracies in prediction of peptide binding. The present backbone library is created by superposing the backbones of all peptides bound to MHC Class II molecules found within the Protein Data Bank and noting the root mean square (RMS) deviation between the Cα atoms of each of the eleven amino-acids located within the binding groove. While this library can be derived from a small number of suitable available mouse and human structures (currently 13), in order to allow for the possibility of even greater variability, the RMS figure for each C"-α position is increased by 50%. The average Cα position of each amino-acid is then determined and a sphere drawn around this point whose radius equals the RMS deviation at that position plus 50%. This sphere represents all allowed Cα positions.

Working from the Cα with the least RMS deviation (that of the amino-acid in Pocket 1 as mentioned above, equivalent to Position 2 of the 11 residues in the binding groove), the sphere is three-dimensionally gridded, and each vertex within the grid is then used as a possible location for a Cα of that amino-acid. The subsequent amide plane, corresponding to the peptide bond to the subsequent amino-acid is grafted onto each of these Cαs and the ϕ and ψ angles are rotated step-wise at set intervals in order to position the subsequent Cα. If the subsequent Cα falls within the 'sphere of allowed positions' for this Cα than the orientation of the dipeptide is accepted, whereas if it falls outside the sphere then the dipeptide is rejected. This process is then repeated for each of the subsequent Cα positions, such that the peptide grows from the Pocket 1 Cα 'seed', until all nine subsequent Cαs have been positioned from all possible permutations of the preceding Cαs. The process is then repeated once more for the single Cα preceding pocket 1 to create a library of backbone Cα positions located within the binding groove.

The number of backbones generated is dependent upon several factors: The size of the "spheres of allowed positions"; the fineness of the gridding of the "primary sphere" at the

Pocket 1 position; the fineness of the step-wise rotation of the ϕ and ψ angles used to position subsequent Cαs. Using this process, a large library of backbones can be created. The larger the backbone library, the more likely it will be that the optimum fit will be found for a particular peptide within the binding groove of an MHC Class II molecule. Inasmuch as all backbones will not be suitable for docking with all the models of MHC Class II molecules due to clashes with amino-acids of the binding domains, for each allele a subset of the library is created comprising backbones which can be accommodated by that allele. The use of the backbone library, in conjunction with the models of MHC Class II molecules creates an exhaustive database consisting of allowed side chain conformations for each amino-acid in each position of the binding groove for each MHC Class II molecule docked with each allowed backbone. This data set is generated using a simple steric overlap function where a MHC Class II molecule is docked with a backbone and an amino-acid side chain is grafted onto the backbone at the desired position. Each of the rotatable bonds of the side chain is rotated step-wise at set intervals and the resultant positions of the atoms dependent upon that bond noted. The interaction of the atom with atoms of side-chains of the binding groove is noted and positions are either accepted or rejected according to the following criteria: The sum total of the overlap of all atoms so far positioned must not exceed a pre-determined value. Thus the stringency of the conformational search is a function of the interval used in the step-wise rotation of the bond and the pre-determined limit for the total overlap. This latter value can be small if it is known that a particular pocket is rigid, however the stringency can be relaxed if the positions of pocket side-chains are known to be relatively flexible. Thus allowances can be made to imitate variations in flexibility within pockets of the binding groove. This conformational search is then repeated for every amino-acid at every position of each backbone when docked with each of the MHC Class II molecules to create the exhaustive database of side-chain conformations.

A suitable mathematical expression is used to estimate the energy of binding between models of MHC Class II molecules in conjunction with peptide ligand conformations which are empirically derived by scanning the large database of backbone/side-chain conformations described above. Thus a protein is scanned for potential T-cell epitopes by subjecting each possible peptide of length varying between 9 and 20 amino-acids (although the length is kept constant for each scan) to the following computations: an MHC Class II molecule is selected together with a peptide backbone allowed for that molecule and the side-chains corresponding to the desired peptide sequence are grafted on. Atom identity and interatomic distance data

relating to a particular side-chain at a particular position on the backbone are collected for each allowed conformation of that amino-acid (obtained from the database described above). This is repeated for each side-chain along the backbone and peptide scores derived using a scoring function. The best score for that backbone is retained and the process repeated for each allowed backbone for the selected model. The scores from all allowed backbones are compared and the highest score is deemed to be the peptide score for the desired peptide in that MHC Class II model. This process is then repeated for each model with every possible peptide derived from the protein being scanned, and the scores for peptides versus models are displayed.

In the context of the present invention, each ligand presented for the binding affinity calculation is an amino-acid segment selected from a peptide or protein as discussed above. Thus, the ligand is a selected stretch of amino acids about 9 to 20 amino acids in length derived from a peptide, polypeptide or protein of known sequence. The terms "amino acids" and "residues" are hereinafter regarded as equivalent terms. The ligand, in the form of the consecutive amino acids of the peptide to be examined grafted onto a backbone from the backbone library, is positioned in the binding cleft of an MHC Class II molecule from the MHC Class II molecule model library via the coordinates of the C"-α atoms of the peptide backbone and an allowed conformation for each side-chain is selected from the database of allowed conformations. The relevant atom identities and interatomic distances are also retrieved from this database and used to calculate the peptide binding score. Ligands with a high binding affinity for the MHC Class II binding pocket are flagged as candidates for site-directed mutagenesis. Amino-acid substitutions are made in the flagged ligand (and hence in the protein of interest) which is then retested using the scoring function in order to determine changes which reduce the binding affinity below a predetermined threshold value. These changes can then be incorporated into the protein of interest to remove T-cell epitopes. Binding between the peptide ligand and the binding groove of MHC Class II molecules involves non-covalent interactions including, but not limited to: hydrogen bonds, electrostatic interactions, hydrophobic (lipophilic) interactions and Van der Waals interactions. These are included in the peptide scoring function as described in detail below. It should be understood that a hydrogen bond is a non-covalent bond which can be formed between polar or charged groups and consists of a hydrogen atom shared by two other atoms. The hydrogen of the hydrogen donor has a positive charge where the hydrogen acceptor has a partial negative charge. For the purposes of peptide/protein interactions, hydrogen bond donors may be either

nitrogens with hydrogen attached or hydrogens attached to oxygen or nitrogen. Hydrogen bond acceptor atoms may be oxygens not attached to hydrogen, nitrogens with no hydrogens attached and one or two connections, or sulphurs with only one connection. Certain atoms, such as oxygens attached to hydrogens or imine nitrogens (e.g. C=NH) may be both hydrogen acceptors or donors. Hydrogen bond energies range from 3 to 7 Kcal/mol and are much stronger than Van der Waal's bonds, but weaker than covalent bonds. Hydrogen bonds are also highly directional and are at their strongest when the donor atom, hydrogen atom and acceptor atom are co-linear. Electrostatic bonds are formed between oppositely charged ion pairs and the strength of the interaction is inversely proportional to the square of the distance between the atoms according to Coulomb's law. The optimal distance between ion pairs is about 2.8Å. In protein/peptide interactions, electrostatic bonds may be formed between arginine, histidine or lysine and aspartate or glutamate. The strength of the bond will depend upon the pKa of the ionizing group and the dielectric constant of the medium although they are approximately similar in strength to hydrogen bonds.

Lipophilic interactions are favorable hydrophobic-hydrophobic contacts that occur between he protein and peptide ligand. Usually, these will occur between hydrophobic amino acid side chains of the peptide buried within the pockets of the binding groove such that they are not exposed to solvent. Exposure of the hydrophobic residues to solvent is highly unfavorable since the surrounding solvent molecules are forced to hydrogen bond with each other forming cage-like clathrate structures. The resultant decrease in entropy is highly unfavorable. Lipophilic atoms may be sulphurs that are neither polar nor hydrogen acceptors and carbon atoms that are not polar.

Van der Waal's bonds are non-specific forces found between atoms which are 3-4Å apart. They are weaker and less specific than hydrogen and electrostatic bonds. The distribution of electronic charge around an atom changes with time and, at any instant, the charge distribution is not symmetric. This transient asymmetry in electronic charge induces a similar asymmetry in neighboring atoms. The resultant attractive forces between atoms reaches a maximum at the Van der Waal's contact distance but diminishes very rapidly at about 1Å to about 2Å. Conversely, as atoms become separated by less than the contact distance, increasingly strong repulsive forces become dominant as the outer electron clouds of the atoms overlap. Although the attractive forces are relatively weak compared to electrostatic and hydrogen bonds (about 0.6 Kcal/mol), the repulsive forces in particular may be very important in determining whether a peptide ligand may bind successfully to a protein.

In one embodiment, the Böhm scoring function (SCORE1 approach) is used to estimate the binding constant. (Böhm, H.J., *J. Comput Aided Mol. Des.*, 8(3):243-256 (1994) which is hereby incorporated in its entirety). In another embodiment, the scoring function (SCORE2 approach) is used to estimate the binding affinities as an indicator of a ligand containing a T-cell epitope (Böhm, H.J., *J. Comput Aided Mol. Des.*, 12(4):309-323 (1998) which is hereby incorporated in its entirety). However, the Böhm scoring functions as described in the above references are used to estimate the binding affinity of a ligand to a protein where it is already known that the ligand successfully binds to the protein and the protein/ligand complex has had its structure solved, the solved structure being present in the Protein Data Bank ("PDB"). Therefore, the scoring function has been developed with the benefit of known positive binding data. To allow for discrimination between positive and negative binders, a repulsion term can optionally be added to the equation. In addition, a more satisfactory estimate of binding energy is achieved by computing the lipophilic interactions in a pairwise manner rather than using the area based energy term of the above Böhm functions. Therefore, in a preferred embodiment, the binding energy is estimated using a modified Böhm scoring function. In the modified Böhm scoring function, the binding energy between protein and ligand (ΔG_{bind}) is estimated considering the following parameters: The reduction of binding energy due to the overall loss of translational and rotational entropy of the ligand (ΔG_0); contributions from ideal hydrogen bonds (ΔG_{hb}) where at least one partner is neutral; contributions from unperturbed ionic interactions (ΔG_{ionic}); lipophilic interactions between lipophilic ligand atoms and lipophilic acceptor atoms (ΔG_{lipo}); the loss of binding energy due to the freezing of internal degrees of freedom in the ligand, i.e., the freedom of rotation about each C-C bond is reduced (ΔG_{rot}); the energy of the interaction between the protein and ligand (E_{vdw}). Consideration of these terms gives equation 1:

$$(\Delta G_{bind})=(\Delta G_0)+(\Delta G_{hb} x N_{hb})+(\Delta G_{ionic} x N_{ionic})+(\Delta G_{lipo} x N_{lipo})+(\Delta G_{rot}+N_{rot})+(E_{vdw}).$$

Where N is the number of qualifying interactions for a specific term and, in one embodiment, ΔG_0, ΔG_{hb}, ΔG_{ionic}, ΔG_{lipo} and ΔG_{rot} are constants which are given the values: 5.4, -4.7, -4.7, -0.17, and 1.4, respectively.

The term N_{hb} is calculated according to equation 2:

$$N_{hb} = \Sigma_{h\text{-}bonds} f(\Delta R, \Delta\alpha) \times f(N_{neighb}) \times f_{pcs}$$

$f(\Delta R, \Delta\alpha)$ is a penalty function which accounts for large deviations of hydrogen bonds from ideality and is calculated according to equation 3:

$f(\Delta R, \Delta\text{-}\square) = f1(\Delta R) \times f2(\Delta\alpha)$

Where: $f1(\Delta R) = 1$ if $\Delta R <= TOL$

or $= 1 - (\Delta R - TOL)/0.4$ if $\Delta R <= 0.4 + TOL$

or $= 0$ if $\Delta R > 0.4 + TOL$

And: $f2(\Delta\alpha) = 1$ if $\Delta\alpha < 30°$

or $= 1-(\Delta\alpha - 30)/50$ if $\Delta\alpha <= 80°$

or $= 0$ if $\Delta\alpha > 80°$

TOL is the tolerated deviation in hydrogen bond length = 0.25Å

ΔR is the deviation of the H-O/N hydrogen bond length from the ideal value = 1.9Å

$\Delta\alpha$ is the deviation of the hydrogen bond angle $\angle_{N/O\text{-}H..O/N}$ from its idealized value of 180°

$f(N_{neighb})$ distinguishes between concave and convex parts of a protein surface and therefore assigns greater weight to polar interactions found in pockets rather than those found at the protein surface.

This function is calculated according to equation 4 below:

$f(N_{neighb}) = (N_{neighb}/N_{neighb,0})^{\alpha}$ where $\alpha = 0.5$

N_{neighb} is the number of non-hydrogen protein atoms that are closer than 5Å to any given protein atom.

$N_{neighb,0}$ is a constant = 25

f_{pcs} is a function which allows for the polar contact surface area per hydrogen bond and therefore distinguishes between strong and weak hydrogen bonds and its value is determined according to the following criteria:

$f_{pcs} = ß$ when $A_{polar}/N_{HB} < 10$ Å^2

or $f_{pcs} = 1$ when $A_{polar}/N_{HB} > 10$ Å^2

A_{polar} is the size of the polar protein-ligand contact surface

N_{HB} is the number of hydrogen bonds

ß is a constant whose value = 1.2

For the implementation of the modified Böhm scoring function, the contributions from ionic interactions, ΔG_{ionic}, are computed in a similar fashion to those from hydrogen bonds described above since the same geometry dependency is assumed.

The term N_{lipo} is calculated according to equation 5 below:

$N_{lipo} = \sum_{iL} f(r_{iL})$

$f(r_{lL})$ is calculated for all lipophilic ligand atoms, l, and all lipophilic protein atoms, L, according to the following criteria:

$f(r_{lL}) = 1$ when $r_{lL} <= R1$ $f(r_{lL}) = (r_{lL} - R1)/(R2-R1)$ when $R2 < r_{lL} > R1$

$f(r_{lL}) = 0$ when $r_{lL} >= R2$

Where: $R1 = r_l^{vdw} + r_L^{vdw} + 0.5$

and $R2 = R1 + 3.0$

and r_l^{vdw} is the Van der Waal's radius of atom l

and r_L^{vdw} is the Van der Waal's radius of atom L

The term N_{rot} is the number of rotable bonds of the amino acid side chain and is taken to be the number of acyclic sp^3 - sp^3 and sp^3 - sp^2 bonds. Rotations of terminal $-CH_3$ or $-NH_3$ are not taken into account.

The final term, E_{VdW}, is calculated according to equation 6 below:

$E_{VdW} = \varepsilon_1\varepsilon_2((r_1^{vdw} + r_2^{vdw})^{12}/r^{12} - (r_1^{vdw} + r_2^{vdw})^6/r^6)$, where:

ε_1 and ε_2 are constants dependant upon atom identity

$r_1^{vdw} + r_2^{vdw}$ are the Van der Waal's atomic radii

r is the distance between a pair of atoms.

With regard to equation 6, in one embodiment, the constants ε_1 and ε_2 are given the atom values: C: 0.245, N: 0.283, O: 0.316, S: 0.316, respectively (i.e. for atoms of Carbon, Nitrogen, Oxygen and Sulphur, respectively). With regards to equations 5 and 6, the Van der Waal's radii are given the atom values C: 1.85, N: 1.75, O: 1.60, S: 2.00Å.

It should be understood that all predetermined values and constants given in the equations above are determined within the constraints of current understandings of protein ligand interactions with particular regard to the type of computation being undertaken herein. Therefore, it is possible that, as this scoring function is refined further as a result of progress in the field of modeling of molecular interactions, these values and constants may change hence any suitable numerical value that gives the desired results in terms of estimating the binding energy of a protein to a ligand may be used and thus fall within the scope of the present invention.

As described above, the scoring function is applied to data extracted from the database of side-chain conformations, atom identities, and interatomic distances. For the purposes of the present description, the number of MHC Class II molecules included in this database is 42 models plus four solved structures. It should be apparent from the above descriptions that the

modular nature of the construction of the computational method of the present invention means that new models can simply be added and scanned with the peptide backbone library and side-chain conformational search function to create additional data sets which can be processed by the peptide scoring function as described above. This allows for the repertoire of scanned MHC Class II molecules to easily be increased, or structures and associated data to be replaced if data are available to create more accurate models of the existing alleles.

It should be understood that, although the above scoring function is relatively simple compared to some sophisticated methodologies that are available, the calculations are performed extremely rapidly. It should also be understood that the objective is not to calculate the true binding energy *per se* for each peptide docked in the binding groove of a selected MHC Class II protein. The underlying objective is to obtain comparative binding energy data as an aid to predicting the location of T-cell epitopes based on the primary structure (i.e. amino acid sequence) of a selected protein. A relatively high binding energy or a binding energy above a selected threshold value would suggest the presence of a T-cell epitope in the ligand. The ligand may then be subjected to at least one round of amino-acid substitution and the binding energy recalculated. Due to the rapid nature of the calculations, these manipulations of the peptide sequence can be performed interactively within the program's user interface on cost-effectively available computer hardware. Major investment in computer hardware is thus not required.

It would be apparent to one skilled in the art that other available software could be used for the same purposes. In particular, more sophisticated software which is capable of docking ligands into protein binding-sites may be used in conjunction with energy minimization. Examples of docking software are: DOCK (Kuntz *et al.*, *J. Mol. Biol.*, 161:269-288 (1982)), LUDI (Böhm, H.J., *J. Comput Aided Mol. Des.*, 8:623-632 (1994)) and FLEXX (Rarey M., *et al.*, *ISMB*, 3:300-308 (1995)). Examples of molecular modeling and manipulation software include: AMBER (Tripos) and CHARMm (Molecular Simulations Inc.). The use of these computational methods would severely limit the throughput of the method of this invention due to the lengths of processing time required to make the necessary calculations. However, it is feasible that such methods could be used as a 'secondary screen' to obtain more accurate calculations of binding energy for peptides which are found to be 'positive binders' via the method of the present invention.

The limitation of processing time for sophisticated molecular mechanic or molecular dynamic calculations is one which is defined both by the design of the software which makes these calculations and the current technology limitations of computer hardware. It may be anticipated that, in the future, with the writing of more efficient code and the continuing increases in speed of computer processors, it may become feasible to make such calculations within a more manageable time-frame. Further information on energy functions applied to macromolecules and consideration of the various interactions that take place within a folded protein structure can be found in: Brooks, B.R., *et al.*, *J. Comput. Chem.*, 4:187-217 (1983) and further information concerning general protein-ligand interactions can be found in: Dauber-Osguthorpe et al., *Proteins*4(1):31-47(1988), which are incorporated herein by reference in their entirety. Useful background information can also be found, for example, in Fasman, G.D., ed., *Prediction of Protein Structure and the Principles of Protein Conformation,* Plenum Press, New York, ISBN: 0-306 4313-9.

The present prediction method can be calibrated against a data set comprising a large number of peptides whose affinity for various MHC Class II molecules has previously been experimentally determined.

According to a preferred embodiment of the method, any one of the specific prediction methods described herein, or any other computer-based method of predicting peptide-MHC Class II interactions that yields numerical scores for each peptide/MHC Class II pair, is calibrated against a data set comprising a large number of peptides whose affinity for various MHC Class II molecules has previously been experimentally determined. By comparison of calculated versus experimental data, a cut of value can be determined above which it is known that all experimentally determined T-cell epitopes are correctly predicted.

Specifically, the computer-derived numerical score is calculated for each peptide/MHC Class II pair in the data set. The score is calculated such that a higher score represents an increased probability of binding. The lowest computer-based score for a peptide/MHC Class II pair that is found experimentally to bind is taken to be a cutoff. All computer-based scores that are significantly below this cutoff score are considered to represent non-binding peptide/MHC Class II pairs, while computer-based scores above the cutoff represent a potential binding peptide/MHC Class II pair. In general for a given computer-based scoring algorithm, there

will be some peptide/MHC Class II combinations that give scores above the cutoff but that do not actually bind. Thus, this preferred embodiment of the method may generate false-positives, but will never or only rarely generate false negatives.

This cutoff-based embodiment of the method is particularly useful when a goal is to eliminate, by mutation, most or all of the T-cell epitopes from a protein. Specifically, according to a more preferred embodiment of the method of the invention, most or all of the T-cell epitopes are removed from a protein as follows. The protein sequence is scanned by a computer-based algorithm for potential T-cell epitopes. Each potential T-cell epitope is given a score, with increasing scores correlated with higher probability of binding to an MHC Class II. Each peptide segment with a score greater than a cutoff is mutated such that the score of the mutated segment is less than the cutoff. Mutations are preferentially chosen that do not reduce the activity of the protein below an activity necessary for a given purpose. A multiply mutated protein, lacking most or all of its computer-predicted T-cell epitopes, is designed. Such a multiply mutated protein is termed a "DeImmunized protein".

The DeImmunized protein is synthesized by standard methods. For example, an artificial DNA sequence encoding the DeImmunized protein is assembled from synthetic oligonucleotides, ligated into an expression vector and functionally linked to elements promoting expression of the DeImmunized protein. The DeImmunized protein is then purified by standard methods. The resulting DeImmunized protein contains mutated amino acids such that genuine T-cell epitopes are eliminated. In addition, the DeImmunized protein will often contain mutated amino acids in segments that are predicted by an algorithm to be T-cell epitopes, but that are not in fact T-cell epitopes. However, significant deleterious consequences do not result from the mutations in the falsely predicted epitopes, because the mutations are chosen to have little effect on protein activity. Moreover, deleterious consequences do not result from the possible introduction of new B cell epitopes into a protein, because the lack of T-cell epitopes prevents a B cell response to the modified protein.

Application of the above-described methodology to various peptides which may be considered for DeImmunization, for modifications to enhance MHC Class II binding for therapeutic purposes, is exemplified below.

The invention may be applied to any biological molecule having a defined biological and / or pharmocological activity with substantially the same primary amino acid sequences as those disclosed herein and would include therefore molecules derived by genetic engineering means or other processes. The term "biological molecule" is used herein for molecules which have a biological function and cause a biological , pharmacological or pharmaceutical effect or activity. Preferably, biological molecules according to the inventions are peptides, polypeptides, proteins. Hereunder proteins, immunoglobulins are preferred. The invention includes also variants and other modification of a specific polypeptide, protein, fusion protein, immunoglobulin which have in principal the same biological activity and a similar (reduced) immunogenicity. Furthermore fragments of antibodies like sFv, Fab, Fab', F(ab')2 and Fc and biologically effective fragments of proteins are included. Antibodies from human origin or humanized antibodies show *per se* lower or no immunogenicity in humans and have no or a lower number of immunogenic epitopes compared to non-human antibodies. Nevertheless there is also a need for de-immunization of such molecules since some of them have been shown to elicit a significant immune response in humans. Furthermore antigens which elicit a not desired and too strong immune response can be modified according to the method of the invention and result in antigens which have a reduced immunogenicity which is however strong enough for using the antigen e.g. as vaccine.

Some molecules, like leptin, such as identified from other mammalian sources have in common many of the peptide sequences of the present disclosure and have in common many peptide sequences with substantially the same sequence as those of the disclosed listing. Such protein sequences equally therefore fall under the scope of the present invention.

The invention relates to analogues of the biological molecules according to the invention in which substitutions of at least one amino acid residue have been made at positions resulting in a substantial reduction in activity of or elimination of one or more potential T-cell epitopes from the protein.

One or more amino acid substitutions at particular points within any of the potential MHC class II ligands identified in the tables of the examples may result in a molecule with a reduced immunogenic potential when administered as a therapeutic to the human host. Preferably, amino acid substitutions are made at appropriate points within the peptide sequence predicted

to achieve substantial reduction or elimination of the activity of the T-cell epitope. In practice an appropriate point will preferably equate to an amino acid residue binding within one of the hydrophobic pockets provided within the MHC class II binding groove. Amino acid residues in the peptide at positions equating to binding within other pocket regions within the MHC binding cleft are also considered and fall under the scope of the present.

It is understood that single amino acid substitutions within a given potential T-cell epitope are the most preferred route by which the epitope may be eliminated. Combinations of substitution within a single epitope may be contemplated and for example can be particularly appropriate where individually defined epitopes are in overlap with each other. Moreover, amino acid substitutions either singly within a given epitope or in combination within a single epitope may be made at positions not equating to the "pocket residues" with respect to the MHC class II binding groove, but at any point within the peptide sequence. All such substitutions fall within the scope of the present.

Amino acid substitutions other than within the peptides identified above may be contemplated particularly when made in combination with substitution(s) made within a listed peptide. For example a change may be contemplated to restore structure or biological activity of the variant molecule. Such compensatory changes and changes to include deletion or addition of particular amino acid residues from the molecule according to the invention resulting in a variant with desired activity and in combination with changes in any of the disclosed peptides fall under the scope of the present.

In another aspect, the present invention relates to nucleic acids encoding said biological molecules having reduced immunogenicity. Methods for making gene constructs and gene products are well known in the art. In a final aspect the present invention relates to pharmaceutical compositions comprising said biological molecules obtainable by the methods disclosed in the present invention, and methods for therapeutic treatment of humans using the modified molecules and pharmaceutical compositions.

As can be seen from the following examples, the computational methods described herein above provide a very good indicator of where T-cell epitopes are likely to be found in any peptide. This, therefore, allows identification of regions of amino acid residue sequences

which, if altered by one or more amino acid residue changes, have the effect of removing T-cell epitopes and thus enhance the therapeutic value of the peptide. By means of this method biological molecules like peptides, proteins, immunoglobulins and fusion proteins and the like having enhanced properties and pharmacological value can be prepared.

The foregoing description and the examples are intended as illustrative, and are not to be taken as limiting. Still other variants within the spirit and scope of this invention are possible and will readily present themselves to those skilled in the art.

EXAMPLE 1

This example shows the T-cell epitope likelihood profile of the autoantigen glutamic acid decarboxylase isoform (GAD 65; MW: 65.000), which is involved in the development of Type I diabetes. This particular protein could be a potential target for increasing the affinity of T-cell epitopes, and also provides a good example for demonstrating the T-cell epitope likelihood index since it is a relatively long peptide (585 amino acid residues) and, therefore, provides a relatively large sample size for profiling.

Shown in FIGURE 5 is the T-cell epitope likelihood profile for GAD 65. The solid line represents the T-cell epitope index calculated using the computational method shown in FIGURE 1, and the dotted line represents the T-cell epitope index predicted using the computational method shown in FIGURE 3 and 4.

EXAMPLE 2

This example shows the T-cell epitope likelihood profile of erythropoietin (EPO), a 193 amino acid residue long cytokine widely used as an intravenously (IV) administered drug to boost red blood cell counts. This represents a good example of a biologic drug with therapeutic value but which could induce inappropriate or undesirable immune responses, especially with the IV route of administration being used, and which may, therefore, benefit from de-immunization after potential T-cell epitopes therein have been identified.

Shown in FIGURE 6 is the T-cell epitope likelihood profile for EPO. The solid line represents the T-cell epitope index calculated using the computational method shown in FIGURE 1, and the dotted line indicates T-cell epitope index predicted using the computational method shown in FIGURE 3 and 4.

EXAMPLE 3

FIGURES 7 and 8 show the T-cell epitope index for the heavy and light chains of a mouse humanized monoclonal antibody directed against A33 antigen. The latter is a transmembrane glycoprotein expressed on the surface of >95% bowel cancers and, therefore, has potential as an anti-cancer therapeutic.

In FIGURES 7 and 8, the solid line represents the T-cell epitope index calculated using the computational method shown in FIGURE 1, and the dotted line represents the T-cell epitope index predicted using the computational method shown in FIGURE 3 and 4.

EXAMPLE 4 (leptin):

One of these therapeutically valuable molecules is human obesity protein, called "leptin". Leptin is a secreted signaling protein of 146 amino acid residues involved in the homeostatic mechanisms maintaining adipose mass (e.g. WO 00/40615, WO 98/28427, WO 96/05309). The protein (and its antagonists) offers significant therapeutic potential for the treatment of diabetes, high blood pressure and cholesterol metabolism. The protein can be produced by recombinant technologies using a number of different host T-cell types. The amino acid sequence of leptin (depicted as one-letter code) is as follows:

VPIQKVQDDTKTLIKTIVTRINDISHTQSVSSKQKVTGLDFIPGLHPILTLSKMDQTLAVYQQILTSM
PSRNVIQISNDLENLRDLLHVLAFSKSCHLPWASGLETLDSLGGVLEASGYSTEVVALSRLQGSLQDM
LWQLDLSPGC

An amino acid sequence which is part of the sequence of an immunogenically non-modified human obesity protein (leptin) and has a potential MHC class II binding activity is selected from the following group identified according to the method of the invention:

VPIQKVQDDTKTL,	QKVQDDTKTLIKT,	KTLIKTIVTRIND,	TLIKTIVTRINDI,
KTIVTRINDISHT,	TIVTRINDISHTQ,	TRINDISHTQSVS,	NDISHTQSVSSKQ,
QSVSSKQKVTGLD,	SSKQKVTGLDFIP,	QKVTGLDFIPGLH,	TGLDFIPGLHPIL,
LDFIPGLHPILTL,	DFIPGLHPILTLS,	PGLHPILTLSKMD,	GLHPILTLSKMDQ,
HPILTLSKMDQTL,	PILTLSKMDQTLA,	LTLSKMDQTLAVY,	SKMDQTLAVYQQI,
QTLAVYQQILTSM,	LAVYQQILTSMPS,	AVYQQILTSMPSR,	QQILTSMPSRNVI,
QILTSMPSRNVIQ,	TSMPSRNVIQISN,	SRNVIQISNDLEN,	RNVIQISNDLENL,
NVIQISNDLENLR,	IQISNDLENLRDL,	NDLENLRDLLHVL,	LENLRDLLHVLAF,
ENLRDLLHVLAFS,	RDLLHVLAFSKSC,	DLLHVLAFSKSCH,	LHVLAFSKSCHLP,
HVLAFSKSCHLPW,	LAFSKSCHLPWAS,	CHLPWASGLETLD,	SGLETLDSLGGVL,
DSLGGVLEASGYS,	SLGGVLEASGYST,	GGVLEASGYSTEV,	SGYSTEVVALSRL,

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

Substitutions carried out according to the methods of the invention leading to the elimination of potential T-cell epitopes of human leptin (WT = wild type) are:

Residue #	WT residue	Substitutions												
3	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
6	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
13	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
14	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
17	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
18	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
21	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
24	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
30	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
36	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
39	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
41	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
42	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
45	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
48	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
49	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
51	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
54	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
58	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
60	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
61	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
64	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
65	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
68	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
73	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
74	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
76	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
80	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
83	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
86	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
87	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
89	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
90	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
92	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
98	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
100	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
104	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
107	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
110	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
113	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
114	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
119	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
123	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
124	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
126	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
129	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
133	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
136	M	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 5 (Il-1R antagonist):

IL-1 is an important inflammatory and immune modulating cytokine with pleiotropic effects on a variety of tissues but may contribute to the pathology associated with rheumatoid arthritis and other diseases associated with local tissue damage. An IL-1 receptor antagonist able to inhibit the action of IL-1 has been purified and the gene cloned [Eisenburg S.P. et al (1990) *Nature*, 343: 341-346; Carter, D.B. et al (1990) *Nature*, 344: 633-637]. Others have provided IL-1Ra molecules [e.g. US 5,075,222].

The amino acid sequence of Il-1Ra (depicted as one-letter code) is as follows:

RPSGRKSSKMQAFRIWDVNQKTFYLRNNQLVAGYLQGPNVNLEEKIDVVPIEPHALFLGIHGGKMCLSC
VKSGDETRLQLEAVNITDLSENRKQDKRFAFIRSDSGPTTSFESAACPGWFLCTAMEADQPVSLTNMPD
EGVMVTKFYFQEDE

An amino acid sequence which is part of the sequence of an immunogenically non-modified IL-1Ra which has a potential MHC class II binding activity is selected from the following group:

RKSSKMQAFRIWD,	SKMQAFRIWDVNQ,	QAFRIWDVNQKTF,
FRIWDVNQKTFYL,	RIWDVNQKTFYLR,	IWDVNQKTFYLRN,
WDVNQKTFYLRNN,	KTFYLRNNQLVAG,	TFYLRNNQLVAGY,
FYLRNNQLVAGYL,	LRNNQLVAGYLQG,	RNNQLVAGYLQGP,
NQLVAGYLQGPNV,	QLVAGYLQGPNVN,	LVAGYLQGPNVNL,
AGYLQGPNVNLEE,	GYLQGPNVNLEEK,	PNVNLEEKIDVVP,
VNLEEKIDVVPIE,	EKIDVVPIEPHAL,	IDVVPIEPHALFL,
DVVPIEPHALFLG,	VPIEPHALFLGIH,	HALFLGIHGGKMC,
ALFLGIHGGKMCL,	LFLGIHGGKMCLS,	LGIHGGKMCLSCV,
GKMCLSCVKSGDE,	MCLSCVKSGDETR,	SCVKSGDETRLQL,
ETRLQLEAVNITD,	TRLQLEAVNITDL,	LQLEAVNITDLSE,
EAVNITDLSENRK,	VNITDLSENRKQD,	TDLSENRKQDKRF,
ENRKQDKRFAFIR,	KRFAFIRSDSGPT,	FAFIRSDSGPTTS,
AFIRSDSGPTTSF,	TSFESAACPGWFL,	SFESAACPGWFLC,
PGWFLCTAMEADQ,	WFLCTAMEADQPV,	TAMEADQPVSLTN,
QPVSLTNMPDEGV,	VSLTNMPDEGVMV,	TNMPDEGVMVTKF,
PDEGVMVTKFYFQ,	EGVMVTKFYFQED,	GVMVTKFYFQEDE

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

Substitutions leading to the elimination of potential T-cell epitopes are:

Residue #	WT Residue	Substitution												
10	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
13	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
15	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
16	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
18	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
23	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
24	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
25	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
30	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
31	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
34	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
35	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
40	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
42	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
46	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
48	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
49	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
51	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
56	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
57	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
58	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
60	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
65	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
67	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
70	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
78	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
80	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
83	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
85	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
88	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
98	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
100	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
101	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
119	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
120	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
121	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
125	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
131	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
133	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
136	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
141	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
142	M	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 6 (BDNF):

Another therapeutically valuable molecule is "human brain-derived neutrophic factor (BDNF)". BNDF is glycoprotein of the nerve growth factor family of proteins. The mature 119 amino acid glycoprotein is processed from a larger pre-cursor to yield a neutrophic factor

that promotes the survival of neuronal cell populations [Jones K.R. & Reichardt, L.F. (1990) Proc. Natl. Acad. Sci U.S.A. 87: 8060-8064]. Such neuronal cells are all located either in the central nervous system or directly connected to it. Recombinant preparations of BNDF have enabled the therapeutic potential of the protein to be explored for the promotion of nerve regeneration and degenerative disease therapy.

The amino acid sequence of human brain-derived neutrophic factor (BDNF) (depicted as one-letter code) is as follows:

HSDPARRGELSVCDSISEWVTAADKKTAVDMSGGTVTVLEKVPVSKGQLKQYFYETKCNPMGYTKEGCR
GIDKRHWNSQCRTTQSYVRALTMDSKKRIGWRFIRIDTSCVCTLTIKRGR

Others have provided modified BNDF molecules [US, 5,770,577] and approaches towards the commercial production of recombinant BNDF molecules [US, 5,986,070].

An amino acid sequence which is part of the sequence of an immunogenically non-modified human brain-derived neurotrophic factor (BDNF) and has a potential MHC class II binding activity is selected from the following group:

GELSVCDSISEWV, LSVCDSISEWVTA, DSISEWVTAADKK, SEWVTAADKKTAV,
EWVTAADKKTAVD, WVTAADKKTAVDM, KTAVDMSGGTVTV, TAVDMSGGTVTVL,
VDMSGGTVTVLEK, GTVTVLEKVPVSK, VTVLEKVPVSKGQ, TVLEKVPVSKGQL,
EKVPVSKGQLKQY, VPVSKGQLKQYFY, GQLKQYFYETKCN, KQYFYETKCNPMG,
QYFYETKCNPMGY, YFYETKCNPMGYT, NPMGYTKEGCRGI, MGYTKEGCRGIDK,
RGIDKRHWNSQCR, RHWNSQCRTTQSY, HWNSQCRTTQSYV, QSYVRALTMDSKK,
SYVRALTMDSKKR, RALTMDSKKRIGW, LTMDSKKRIGWRF, KRIGWRFIRIDTS,
IGWRFIRIDTSCV, GWRFIRIDTSCVC, WRFIRIDTSCVCT, RFIRIDTSCVCTL,
IRIDTSCVCTLTI, IDTSCVCTLTIKR

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

Substitutions leading to the elimination of potential T-cell epitopes of human brain-derived neutrophic factor (BDNF) (WT = wild type) are:

Residue #	WT Residue	Substitution												
10	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
16	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
20	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
29	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
31	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
36	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
38	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
39	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
42	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
44	V	A	C	D	E	G	H	K	N	P	Q	R	S	T

49	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
52	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
53	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
54	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
61	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
63	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
71	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
76	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
86	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
87	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
90	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
92	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
98	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
100	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
102	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
103	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
105	I	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 7 (EPO)

Another therapeutically valuable molecule is erythropoietin (EPO). EPO is a glycoprotein hormone involved in the maturation of erythroid progenitor cells into erythrocytes. Naturally occurring EPO is produced by the liver during foetal life and by the kidney of adults and circulates in the blood to stimulate production of red blood cells in bone marrow. Anaemia is almost invariably a consequence of renal failure due to decreased production of EPO from the kidney. Recombinant EPO is used as an effective treatment of anaemia resulting from chronic renal failure. Recombinant EPO (expressed in mammalian cells) having the amino acid sequence 1-165 of human erythropoietin [Jacobs, K. et al (1985) Nature, 313: 806-810; Lin, F.-K. et al (1985) Proc. Natl. Acad. Sci. U.S.A. 82:7580-7585] contains three N-linked and one O-linked oligosaccharide chains each containing terminal sialic acid residues. The latter are significant in enabling EPO to evade rapid clearance from the circulation by the hepatic asialoglycoprotein binding protein.

The amino acid sequence of EPO (depicted as one-letter code) is as follows:

APPRLICDSRVLERYLLEAKEAENITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLAL
LSEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRTITADTF
RKLFRVYSNFLRGKLKLYTGEACRTGDR

An amino acid sequence which is part of the sequence of an immunogenically non-modified human erythropoietin (EPO) and has a potential MHC class II binding activity is selected from the following group:

PRLICDSRVLERY, RLICDSRVLERYL, ICDSRVLERYLLE, CDSRVLERYLLEA, SRVLERYLLEAKE,
RVLERYLLEAKEA, LERYLLEAKEAEN, ERYLLEAKEAENI, RYLLEAKEAENIT, YLLEAKEAENITT,
LEAKEAENITTGC, KEAENITTGCAEH, ENITTGCAEHCSL, CSLNENITVPDTK, NENITVPDTKVNF,

ENITVPDTKVNFY, NITVPDTKVNFYA, ITVPDTKVNFYAW, TKVNFYAWKRMEV, VNFYAWKRMEVGQ, NFYAWKRMEVGQQ, YAWKRMEVGQQAV, KRMEVGQQAVEVW, RMEVGQQAVEVWQ, MEVGQQAVEVWQG, QAVEVWQGLALLS, AVEVWQGLALLSE, VEVWQGLALLSEA, EVWQGLALLSEAV, VWQGLALLSEAVL, WQGLALLSEAVLR, QGLALLSEAVLRG, LALLSEAVLRGQA, ALLSEAVLRGQAL, LSEAVLRGQALLV, SEAVLRGQALLVN, EAVLRGQALLVNS, AVLRGQALLVNSS, QALLVNSSQPWEP, ALLVNSSQPWEPL, LLVNSSQPWEPLQ, QPWEPLQLHVDKA, EPLQLHVDKAVSG, LQLHVDKAVSGLR, LHVDKAVSGLRSL, KAVSGLRSLTTLL, SGLRSLTTLLRAL, RSLTTLLRALGAQ, SLTTLLRALGAQK, TTLLRALGAQKEA, TLLRALGAQKEAI, RALGAQKEAISPP, AQKEAISPPDAAS, EAISPPDAASAAP, SPPDAASAAPLRT, ASAAPLRTITADT, APLRTITADTFRK, RTITADTFRKLFR, TITADTFRKLFRV, DTFRKLFRVYSNF, RKLFRVYSNFLRG, KLFRVYSNFLRGK, FRVYSNFLRGKLK, RVYSNFLRGKLKL, YSNFLRGKLKLYT, SNFLRGKLKLYTG, NFLRGKLKLYTGE, RGKLKLYTGEACR, GKLKLYTGEACRT, LKLYTGEACRTGD, KLYTGEACRTGDR

Substitutions leading to the elimination of potential T-cell epitopes of human erythropoietin (EPO) (WT = wild type) are:

Residue #	WT residue	Substitutions											
5	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
6	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
11	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
12	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
15	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
16	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
17	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
25	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
35	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
39	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
41	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
46	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
48	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
49	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
51	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
54	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
56	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
61	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
63	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
64	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
67	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
69	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
70	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
74	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
75	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
80	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
81	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
82	N	A	C	D	E	G	H	K	N	P	Q	R	S	T
88	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
91	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
93	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
95	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
99	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
102	L	A	C	D	E	G	H	K	N	P	Q	R	S	T

Residue #	WT residue	Substitutions											
105	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
108	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
109	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
112	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
119	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
130	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
133	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
138	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
141	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
142	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
144	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
145	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
148	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
149	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
153	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
155	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
156	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 8 (G-CSF)

Granulocyte colony stimulating factor (G-CSF) is an important haemopoietic cytokine currently used in treatment of indications where an increase in blood neutrophils will provide benefits. These include cancer therapy, various infectious diseases and related conditions such as sepsis. G-CSF is also used alone, or in combination with other compounds and cytokines in the *ex vivo* expansion of haemopoeitic cells for bone marrow transplantation.

Two forms of human G-CSF are commonly recognized for this cytokine. One is a protein of 177 amino acids, the other a protein of 174 amino acids [Nagata et al. (1986), EMBO J. 5: 575-581], the 174 amino acid form has been found to have the greatest specific *in vivo* biological activity. Recombinant DNA techniques have enabled the production of commercial scale quantities of G-CSF exploiting both eukaryotic and prokaryotic host cell expression systems.

The amino acid sequence of human granulocyte colony stimulating factor (G-CSF) (depicted as one-letter code) is as follows:

TPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLLGHSLGIPWAPLSSCPSQALQLAGCLS
QLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPTQGAMPAFASAFQRRAGGVL
VASHLQSFLEVSYRVLRHLAQP.

Other polypeptide analogues and peptide fragments of G-CSF have been previously disclosed, including forms modified by site-specific amino acid substitutions and or by modification by chemical adducts. Thus US 4,810,643 discloses analogues with the particular Cys residues

replaced with another amino acid, and G-CSF with an Ala residue in the first (N-terminal) position. EP 0335423 discloses the modification of at least one amino group in a polypeptide having G-CSF activity. EP 0272703 discloses G-CSF derivatives having amino acid substituted or deleted near the N terminus. EP 0459630 discloses G-CSF derivatives in which Cys 17 and Asp 27 are replaced by Ser residues. EP 0 243 153 discloses G-CSF modified by inactivating at least one yeast KEX2 protease processing site for increased yield in recombinant production and US 4,904,584 discloses lysine altered proteins. WO 90/12874 discloses further Cys altered variants and Australian patent document AU 10948/92 discloses the addition of amino acids to either terminus of a G-CSF molecule for the purpose of aiding in the folding of the molecule after prokaryotic expression. AU-76380/91, discloses G-CSF variants at positions 50-56 of the G-CSF 174 amino acid form, and positions 53-59 of the 177 amino acid form. Additional changes at particular His residues were also disclosed.

An amino acid sequence which is part of the sequence of an immunogenically non-modified human granulocyte colony stimulating factor (G-CSF) and has a potential MHC class II binding activity is selected from the following group:

TPLGPASSLPQSF, SSLPQSFLLKCLE, QSFLLKCLEQVRK, SFLLKCLEQVRKI,
FLLKCLEQVRKIQ, KCLEQVRKIQGDG, EQVRKIQGDGAAL, RKIQGDGAALQEK,
AALQEKLVSECAT, EKLVSECATYKLC, KLVSECATYKLCH, AALQEKLCATYKL,
EKLCATYKLCHPE, ATYKLCHPEELVL, YKLCHPEELVLLG, EELVLLGHSLGIP,
ELVLLGHSLGIPW, HSLGIPWAPLSSC, IPWAPLSSCPSQA, APLSSCPSQALQL,
QALQLAGCLSQLH, GCLSQLHSGLFLY, SQLHSGLFLYQGL, SGLFLYQGLLQAL,
GLFLYQGLLQALE, LFLYQGLLQALEG, FLYQGLLQALEGI, QGLLQALEGISPE,
GLLQALEGISPEL, QALEGISPELGPT, EGISPELGPTLDT, PTLDTLQLDVADF,
DTLQLDVADFATT, LQLDVADFATTIW, LDVADFATTIWQQ, TTIWQQMEELGMA,
TIWQQMEELGMAP, QQMEELGMAPALQ, EELGMAPALQPTQ, LGMAPALQPTQGA,
PALQPTQGAMPAF, GAMPAFASAFQRR, PAFASAFQRRAGG, SAFQRRAGGVLVA,
GGVLVASHLQSFL, GVLVASHLQSFLE, VLVASHLQSFLEV, SHLQSFLEVSYRV,
QSFLEVSYRVLRH, SFLEVSYRVLRHL, LEVSYRVLRHLAQ

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

Substitutions leading to the elimination of potential T-cell epitopes of human granulocyte colony stimulating factor (G-CSF) (WT = wild type) are:

Residue #	WT Residue	Substitution												
3	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
9	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
14	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
15	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
18	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
21	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
24	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
31	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
35	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
39	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
41	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
47	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
48	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
49	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
50	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
54	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
56	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
58	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
61	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
69	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
71	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
75	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
78	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
82	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
83	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
84	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
85	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
88	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
89	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
92	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
95	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
99	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
103	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
106	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
108	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
110	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
113	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
117	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
118	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
121	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
124	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
130	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
137	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
140	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
144	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
151	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
152	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
153	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
157	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
160	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
161	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
163	V	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 9 (KGF):

Another valuable molecule is keratinocyte growth factor (KGF). KGF is a member of the fibroblast growth factor (FGF) / heparin-binding growth factor family of proteins. It is a secreted glycoprotein expressed predominantly in the lung, promoting wound healing by stimulating the growth of keratinocytes and other epithelial cells [Finch et al (1989), Science 24: 752-755; Rubin et al (1989), Proc. Natl. Acad. Sci. U.S.A. 86: 802-806]. The mature (processed) form of the glycoprotein comprises 163 amino acid residues and may be isolated from conditioned media following culture of particular cell lines [Rubin et al, (1989) *ibid.*], or produced using recombinant techniques [Ron et al (1993) J. Biol. Chem. 268: 2984-2988]. The protein is of therapeutic value for the stimulation of epithelial cell growth in a number of significant disease and injury repair settings. This disclosure specifically pertains the human KGF protein being the mature (processed) form of 163 amino acid residues.

Others have also provided KGF molecules [e.g. US, 6,008,328; WO90/08771;] including modified KGF [Ron et al (1993) *ibid*; WO9501434]. However, such teachings have not recognized the importance of T-cell epitopes to the immunogenic properties of the protein nor have been conceived to directly influence said properties in a specific and controlled way according to the scheme of the present invention.

The amino acid sequence of keratinocyte growth factor (KGF) (depicted as one-letter code) is as follows:

MCNDMTPEQMATNVNCSSPERHTRSYDYMEGGDIRVRRLFCRTQWYLRIDKRGKVKGTQEMKNNYNIME
IRTVAVGIVAIKGVESEFYLAMNKEGKLYAKKECNEDCNFKELILENHYNTYASAKWTHNGGEMFVALN
QKGIPVRGKKTKKEQKTAHFLPMAIT

An amino acid sequence which is part of the sequence of an immunogenically non-modified human keratinocyte growth factor (KGF) and has a potential MHC class II binding activity is selected from the following group:

NDMTPEQMATNVN,	DMTPEQMATNVNC,	EQMATNVNCSSPE,	TNVNCSSPERHTR,
RSYDYMEGGDIRV,	YDYMEGGDIRVRR,	DYMEGGDIRVRRL,	GDIRVRRLFCRTQ,
IRVRRLFCRTQWY,	RRLFCRTQWYLRI,	RLFCRTQWYLRID,	TQWYLRIDKRGKV,
QWYLRIDKRGKVK,	WYLRIDKRGKVKG,	LRIDKRGKVKGTQ,	GKVKGTQEMKNNY,
QEMKNNYNIMEIR,	NNYNIMEIRTVAV,	YNIMEIRTVAVGI,	NIMEIRTVAVGIV,
MEIRTVAVGIVAI,	RTVAVGIVAIKGV,	VAVGIVAIKGVES,	VGIVAIKGVESEF,
VAIKGVESEFYLA,	KGVESEFYLAMNK,	SEFYLAMNKEGKL,	EFYLAMNKEGKLY,
FYLAMNKEGKLYA,	LAMNKEGKLYAKK,	GKLYAKKECNEDC,	KLYAKKECNEDCN,
CNFKELILENHYN,	KELILENHYNTYA,	ELILENHYNTYAS,	LILENHYNTYASA,

NHYNTYASAKWTH, NTYASAKWTHNGG, AKWTHNGGEMFVA, GEMFVALNQKGIP, EMFVALNQKGIPV, FVALNQKGIPVRG, VALNQKGIPVRGK, KGIPVRGKKTKKE, IPVRGKKTKKEQK, KTKKEQKTAHFLP

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

Substitutions leading to the elimination of potential T-cell epitopes of human keratinocyte growth factor (KGF) (WT = wild type) are:

Residue #	WT residue	Substitution												
5	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
10	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
14	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
26	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
28	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
29	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
34	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
36	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
39	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
40	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
45	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
46	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
47	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
49	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
55	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
61	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
65	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
67	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
68	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
70	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
73	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
75	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
77	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
78	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
80	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
83	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
87	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
88	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
89	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
91	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
97	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
98	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
109	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
112	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
113	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
114	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
118	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
121	Y	A	C	D	E	G	H	K	N	P	Q	R	S	T
126	W	A	C	D	E	G	H	K	N	P	Q	R	S	T
133	M	A	C	D	E	G	H	K	N	P	Q	R	S	T
134	F	A	C	D	E	G	H	K	N	P	Q	R	S	T
135	V	A	C	D	E	G	H	K	N	P	Q	R	S	T
137	L	A	C	D	E	G	H	K	N	P	Q	R	S	T
142	I	A	C	D	E	G	H	K	N	P	Q	R	S	T
144	V	A	C	D	E	G	H	K	N	P	Q	R	S	T

EXAMPLE 10 (soluble TNF RI):

The sTNF-RI (soluble tumor necrosis factor receptor type I) is a derivative of the human tumor necrosis factor receptor described previously [Gray, P.W. et al (1990) *Proc. Nat. Acad. Sci. U.S.A.* 87: 7380-7384; Loetschere, H. et al, (1990) *Cell* 61: 351-359; Schall, T.J. et al (1990) *Cell* 61: 361-370], comprising the extracellular domain of the intact receptor and exhibiting an approximate molecular weight of 30KDa. Additional soluble TNF inhibitors and in particular a 40KDa form are also known [US 6,143,866]. The soluble forms are able to bind tumor necrosis factor alpha with high affinity and inhibit the cytotoxic activity of the cytokine *in vitro*. Recombinant preparations of sTNF-RI are of significant therapeutic value for the treatment of diseases where an excess level of tumor necrosis factor is causing a pathogenic effect. Indications such as cachexia, sepsis and autoimmune disorders including, and in particular, rheumatoid arthritis and others may be targeted by such therapeutic preparations of sTNF-RI. Others including Brewer et al., US, 6,143,866, have provided modified sTNF-RI molecules

Peptide sequences in a human 30KDa sTNF-RI with potential human MHC class II binding activity:

DSVCPQGKYIHPQ, KYIHPQNNSICCT, NSICCTKCHKGTY, TYLYNDCPGPGQD,
YLYNDCPGPGQDT, NHLRHCLSCSKCR, HCLSCSKCRKEMG, KEMGQVEISSCTV,
GQVEISSCTVDRD, VEISSCTVDRDTV, CTVDRDTVCGCRK, DTVCGCRKNQYRH,
NQYRHYWSENLFQ, RHYWSENLFQCFN, HYWSENLFQCFNC, ENLFQCFNCSLCL,
NLFQCFNCSLCLN, QCFNCSLCLNGTV, CSLCLNGTVHLSC, LCLNGTVHLSCQE,
GTVHLSCQEKQNT, VHLSCQEKQNTVC, EKQNTVCTCHAGF, NTVCTCHAGFFLR,
GFFLRENECVSCS, FFLRENECVSCSN, ECVSCSNCKKSLE, KSLECTKLCLPQI,
TKLCLPQIENVKG, LCLPQIENVKGTE, PQIENVKGTEDSG, SGTTVLLPLVIFF

Any of the above-cited peptide sequences can be used for modifying by exchanging one or more amino acids to obtain a sequence having a reduced or no immunogenicity .

EXAMPLE 11 (soluble TNF-R2):

Soluble tumor necrosis factor receptor 2 (sTNF-R2) is a derivative of the human tumor necrosis factor receptor 2 described previously [Smith, C.A. et al (1990) *Science* 248: 1019-1023; Kohno, T. et al (1990) *Proc. Nat. Acad. Sci. U.S.A.* 87: 8331-8335; Beltinger, C.P. et al (1996) *Genomics* 35:94-100] comprising the extracellular domain of the intact receptor. The soluble forms are able to bind tumour necrosis factor with high affinity and inhibit the

cytotoxic activity of the cytokine *in vitro*. Recombinant preparations of sTNF-R2 are of significant therapeutic value for the treatment of diseases where an excess level of tumour necrosis factor is causing a pathogenic effect. A particular recombinant preparation termed ethanercept has gained clinical approval for the treatment of rheumatoid arthritis and this and other similar agents may be of value in the treatment of other indications such as cachexia, sepsis and autoimmune disorders. Ethanercept is a dimeric fusion protein comprising the extracellular domain of the human TNFR2 molecule in combination with the Fc domain of the human IgG1 molecule. The dimeric molecule comprises 934 amino acids [US,5,395,760; US,5,605,690; US,5,945,397, US, RE36,755].

Peptide sequences in the TNF binding domain of the human TNFR2 protein with potential human MHC class II binding activity are:

TPYAPEPGSTCRL, CRLREYYDQTAQM, REYYDQTAQMCCS, EYYDQTAQMCCSK,
AQMCCSKCSPGQH, KCSPGQHAKVFCT, AKVFCTKTSDTVC, KVFCTKTSDTVCD,
STYTQLWNWVPEC, TQLWNWVPECLSC, QLWNWVPECLSCG, NWVPECLSCGSRC,
ECLSCGSRCSSDQ, SRCSSDQEVTQAC, QEVTQACTREQNR, QNRICTCRPGWYC,
NRICTCRPGWYCA, PGWYCALSKQEGC, GWYCALSKQEGCR, CALSKQEGCRLCA,
APLRKCRPGFGVA, PGFGVARPGTETS, FGVARPGTETSDV, SDVVCKPCAPGTF,
GTFSNTTSSTDIC, TDICRPHQICNVV, HQICNVVAIPGNA, ICNVVAIPGNASR,
CNVVAIPGNASRD, NVVAIPGNASRDA, VAIPGNASRDAVC, DAVCTSTTTPTRS,
TRSMAPGAVHLPQ, RSMAPGAVHLPQP, VHLPQPVSTRSQH, QPVSTRSQHTQPT,
PEPSTAPSTSFLL, SFLLPMGPSPPAE, FLLPMGPSPPAEG

EAXAMPLE 12 (ß-GCR)

Beta-Glucocerebrosidase (b-D-glucosyl-N-acylsphingosine glucohydrolase, E.C. 3.2.1.45) is a monomeric glycoprotein of 497 amino acid residues. The enzyme catalyses the hydrolysis of the glycolipid glucocerebroside to glucose and ceramide. Deficiency in GCR activity results in a lysosomal storage disease referred to as Gaucher disease. The disease is characterised by the accumulation of glucocerebroside engorged tissue macrophages that accumulate in the liver, spleen, bone marrow and other organs. The disease has varying degrees of severity from type 1 disease with haematologic problems but no neuronal involvement, to type 2 disease manifesting early after birth with extensive neuronal involvement and is universally progressive and fatal within 2 years of age. Type 3 disease is also recognised in some classifications and also shows neurologic involvement. Previously the only useful therapy for Gaucher disease has been administration of GCR derived from human placenta (known as alglucerase) but more recently pharmaceutical preparations of recombinant GCR ("ceredase"

and "cerezyme") have shown efficacy in the treatment of type I disease [Niederau, C. et al (1998) *Eur. J. Med. Res.* 3: 25-30].

Peptide sequences in human GCR with potential human MHC class II binding activity are:

PCIPKSFGYSSVV, KSFGYSSVVCVCN, FGYSSVVCVCNAT, SSVVCVCNATYCD, SVVCVCNATYCDS,
VCVCNATYCDSFD, ATYCDSFDPPTFP, DSFDPPTFPALGT, PTFPALGTFSRYE, PALGTFSRYESTR,
GTFSRYESTRSGR, SRYESTRSGRRME, GRRMELSMGPIQA, RRMELSMGPIQAN, RMELSMGPIQANH,
MELSMGPIQANHT, LSMGPIQANHTGT, MGPIQANHTGTGL, GPIQANHTGTGLL, TGLLLTLQPEQKF,
GLLLTLQPEQKFQ, LLLTLQPEQKFQK, LTLQPEQKFQKVK, TLQPEQKFQKVKG, PEQKFQKVKGFGG,
QKFQKVKGFGGAM, QKVKGFGGAMTDA, KGFGGAMTDAAAL, GFGGAMTDAAALN, GAMTDAAALNILA,
AMTDAAALNILAL, MTDAAALNILALS, AALNILALSPPAQ, ALNILALSPPAQN, LNILALSPPAQNL,
NILALSPPAQNLL, LALSPPAQNLLLK, ALSPPAQNLLLKS, PAQNLLLKSYFSE, AQNLLLKSYFSEE,
QNLLLKSYFSEEG, NLLLKSYFSEEGI, LLLKSYFSEEGIG, KSYFSEEGIGYNI, SYFSEEGIGYNII,
FSEEGIGYNIIRV, EGIGYNIIRVPMA, GIGYNIIRVPMAS, IGYNIIRVPMASC, YNIIRVPMASCDF,
NIIRVPMASCDFS, IIRVPMASCDFSI, IRVPMASCDFSIR, VPMASCDFSIRTY, PMASCDFSIRTYT,
SCDFSIRTYTYAD, CDFSIRTYTYADT, FSIRTYTYADTPD, RTYTYADTPDDFQ, TYTYADTPDDFQL,
YTYADTPDDFQLH, ADTPDDFQLHNFS, PDDFQLHNFSLPE, DDFQLHNFSLPEE, FQLHNFSLPEEDT,
HNFSLPEEDTKLK, FSLPEEDTKLKIP, SLPEEDTKLKIPL, EEDTKLKIPLIHR, TKLKIPLIHRALQ,
KLKIPLIHRALQL, LKIPLIHRALQLA, IPLIHRALQLAQR, PLIHRALQLAQRP, HRALQLAQRPVSL,
RALQLAQRPVSLL, ALQLAQRPVSLLA, LQLAQRPVSLLAS, RPVSLLASPWTSP, PVSLLASPWTSPT,
VSLLASPWTSPTW, SLLASPWTSPTWL, SPWTSPTWLKTNG, TSPTWLKTNGAVN, PTWLKTNGAVNGK,
TWLKTNGAVNGKG, GAVNGKGSLKGQP, GSLKGQPGDIYHQ, GDIYHQTWARYFV, DIYHQTWARYFVK,
QTWARYFVKFLDA, WARYFVKFLDAYA, ARYFVKFLDAYAE, RYFVKFLDAYAEH, YFVKFLDAYAEHK,
FVKFLDAYAEHKL, VKFLDAYAEHKLQ, KFLDAYAEHKLQF, DAYAEHKLQFWAV, YAEHKLQFWAVTA,
HKLQFWAVTAENE, LQFWAVTAENEPS, QFWAVTAENEPSA, FWAVTAENEPSAG, WAVTAENEPSAGL,
VTAENEPSAGLLS, PSAGLLSGYPFQC, AGLLSGYPFQCLG, GLLSGYPFQCLGF, SGYPFQCLGFTPE,
YPFQCLGFTPEHQ, QCLGFTPEHQRDF, LGFTPEHQRDFIA, FTPEHQRDFIARD, RDFIARDLGPTLA,
DFIARDLGPTLAN, RDLGPTLANSTHH, LGPTLANSTHHNV, PTLANSTHHNVRL, HNVRLLMLDDQRL,
VRLLMLDDQRLLL, RLLMLDDQRLLLP, LLMLDDQRLLLPH, LMLDDQRLLLPHW, DDQRLLLPHWAKV,
DQRLLLPHWAKVV, QRLLLPHWAKVVL, RLLLPHWAKVVLT, LLLPHWAKVVLTD, PHWAKVVLTDPEA,
WAKVVLTDPEAAK, AKVVLTDPEAAKY, KVVLTDPEAAKYV, VVLTDPEAAKYVH, EAAKYVHGIAVHW,
AKYVHGIAVHWYL, KYVHGIAVHWYLD, YVHGIAVHWYLDF, HGIAVHWYLDFLA, IAVHWYLDFLAPA,
VHWYLDFLAPAKA, HWYLDFLAPAKAT, WYLDFLAPAKATL, LDFLAPAKATLGE, DFLAPAKATLGET,
AKATLGETHRLFP, ATLGETHRLFPNT, GETHRLFPNTMLF, ETHRLFPNTMLFA, THRLFPNTMLFAS,
HRLFPNTMLFASE, RLFPNTMLFASEA, FPNTMLFASEACV, NTMLFASEACVGS, TMLFASEACVGSK,
MLFASEACVGSKF, ACVGSKFWEQSVR, GSKFWEQSVRLGS, SKFWEQSVRLGSW, KFWEQSVRLGSWD,
QSVRLGSWDRGMQ, VRLGSWDRGMQYS, RLGSWDRGMQYSH, GSWDRGMQYSHSI, WDRGMQYSHSIIT,
RGMQYSHSIITNL, MQYSHSIITNLLY, QYSHSIITNLLYH, YSHSIITNLLYHV, HSIITNLLYHVVG,
SIITNLLYHVVGW, TNLLYHVVGWTDW, NLLYHVVGWTDWN, LLYHVVGWTDWNL, YHVVGWTDWNLAL,
HVVGWTDWNLALN, VVGWTDWNLALNP, VGWTDWNLALNPE, TDWNLALNPEGGP, WNLALNPEGGPNW,
LALNPEGGPNWVR, PNWVRNFVDSPII, NWVRNFVDSPIIV, RNFVDSPIIVDIT, NFVDSPIIVDITK,

SPIIVDITKDTFY, PIIVDITKDTFYK, IIVDITKDTFYKQ, VDITKDTFYKQPM, DTFYKQPMFYHLG, TFYKQPMFYHLGH, QPMFYHLGHFSKF, PMFYHLGHFSKFI, MFYHLGHFSKFIP, YHLGHFSKFIPEG, GHFSKFIPEGSQR, SKFIPEGSQRVGL, KFIPEGSQRVGLV, IPEGSQRVGLVAS, QRVGLVASQKNDL, VGLVASQKNDLDA, GLVASQKNDLDAV, SQKNDLDAVALMH, NDLDAVALMHPDG, DAVALMHPDGSAV, VALMHPDGSAVVV, ALMHPDGSAVVVV, SAVVVVLNRSSKD, AVVVVLNRSSKDV, VVVVLNRSSKDVP, VVVLNRSSKDVPL, VVLNRSSKDVPLT, KDVPLTIKDPAVG, VPLTIKDPAVGFL, PLTIKDPAVGFLE, LTIKDPAVGFLET, PAVGFLETISPGY, VGFLETISPGYSI, GFLETISPGYSIH, FLETISPGYSIHT, ETISPGYSIHTYL, PGYSIHTYLWHRQ, PGYSIHTYLWRRQ

EXAMPLE 13 (Protein C):

Protein C is a vitamin K dependent serine-protease involved in the regulation of blood coagulation. The protein is activated by thrombin to produce activated protein C which in turn degrades (down regulates) Factors Va and VIIIa in the coagulation cascade. Protein C is expressed in the liver as a single chain precursor and undergoes a series of processing events resulting in a molecule comprising a light chain and a heavy chain held together by di-sulphide linkage. Protein C is activated by cleavage of a tetradecapeptide from the N-terminus of the heavy chain by thrombin. Pharmaceutical preparations of protein C in native or activated form, have value in the treatment of patients with vascular disorders and or acquired deficiencies in protein C. Such patients include therefore individuals suffering from thrombotic stroke, or protein C deficiency associated with sepsis, transplantation procedures, preganacy, severe burns, major surgery or other severe traumas. Protein C is also used in the treatment of individuals with hereditary protein C deficiency. This disclosure specifically pertains the human protein C being the mature (processed) form comprising a light chain of 155 amino acid residues and a heavy chain of 262 amino acid residues [Foster, D.C. et al (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82: 4673-4677; Beckman, R.J. et al (1985) *Nucleic Acids Res.* 13: 5233-5247]. Others have provided protein C molecules including activated protein C formulations and methods of use [US,6159,468; US,6,156,734; US,6,037,322; US,5,618,714]. Peptide sequences in human protein C heavy-chain with potential human MHC class II binding activity are:

DQEDQVDPRLIDG, QEDQVDPRLIDGK, DQVDPRLIDGKMT, QVDPRLIDGKMTR, VDPRLIDGKMTRR, DPRLIDGKMTRRG, PRLIDGKMTRRGD, RLIDGKMTRRGDS, SPWQVVLLDSKKK, WQVVLLDSKKKLA, QVVLLDSKKKLAC, VVLLDSKKKLACG, VLLDSKKKLACGA, DSKKKLACGAVLI, SKKKLACGAVLIH, KKLACGAVLIHPS, CGAVLIHPSWVLT, GAVLIHPSWVLTA, VLIHPSWVLTAAH, PSWVLTAAHCMDE, SWVLTAAHCMDES, WVLTAAHCMDESK, AAHCMDESKKLLV, HCMDESKKLLVRL, SKKLLVRLGEYDL, KKLLVRLGEYDLR, KLLVRLGEYDLRR, LLVRLGEYDLRRW, VRLGEYDLRRWEK, RLGEYDLRRWEKW, LGEYDLRRWEKWE, GEYDLRRWEKWEL, YDLRRWEKWELDL, RRWEKWELDLDIK, EKWELDLDIKEVF,

WELDLDIKEVFVH, LDLDIKEVFVHPN, LDIKEVFVHPNYS, IKEVFVHPNYSKS, KEVFVHPNYSKST, EVFVHPNYSKSTT, VFVHPNYSKSTTD, PNYSKSTTDNDIA, SKSTTDNDIALLH, TTDNDIALLHLAQ, TDNDIALLHLAQP, DNDIALLHLAQPA, NDIALLHLAQPAT, IALLHLAQPATLS, ALLHLAQPATLSQ, LHLAQPATLSQTI, AQPATLSQTIVPI, PATLSQTIVPICL, ATLSQTIVPICLP, TLSQTIVPICLPD, QTIVPICLPDSGL, TIVPICLPDSGLA, IVPICLPDSGLAE, VPICLPDSGLAER, ICLPDSGLAEREL, PDSGLAERELNQA, SGLAERELNQAGQ, GLAERELNQAGQE, LAERELNQAGQET, RELNQAGQETLVT, GQETLVTGWGYHS, ETLVTGWGYHSSR, TLVTGWGYHSSRE, TGWGYHSSREKEA, WGYHSSREKEAKR, WGYHSSREKEAKR, SREKEAKRNRTFV, RNRTFVLNFIKIP, NRTFVLNFIKIPV, RTFVLNFIKIPVV, TFVLNFIKIPVVP, FVLNFIKIPVVPH, LNFIKIPVVPHNE, NFIKIPVVPHNEC, IKIPVVPHNECSE, IPVVPHNECSEVM, PVVPHNECSEVMS, VVPHNECSEVMSN, NECSEVMSNMVSE, SEVMSNMVSENML, EVMSNMVSENMLC, VMSNMVSENMLCA, SNMVSENMLCAGI, MVSENMLCAGILG, VSENMLCAGILGD, ENMLCAGILGDRQ, NMLCAGILGDRQD, AGILGDRQDACEG, GILGDRQDACEGD, GPMVASFHGTWFL, PMVASFHGTWFLV, MVASFHGTWFLVG, ASFHGTWFLVGLV, GTWFLVGLVSWGE, TWFLVGLVSWGEG, WFLVGLVSWGEGC, FLVGLVSWGEGCG, VGLVSWGEGCGLL, GLVSWGEGCGLLH, VSWGEGCGLLHNY, EGCGLLHNYGVYT, CGLLHNYGVYTKV, GLLHNYGVYTKVS, LHNYGVYTKVSRY, HNYGVYTKVSRYL, YGVYTKVSRYLDW, GVYTKVSRYLDWI, TKVSRYLDWIHGH, SRYLDWIHGHIRD, RYLDWIHGHIRDK, LDWIHGHIRDKEA, DWIHGHIRDKEAP, HGHIRDKEAPQKS, GHIRDKEAPQKSW

Peptide sequences in human protein C light-chain with potential human MHC class II binding activity are:

NSFLEELRHSSLE, SFLEELRHSSLER, EELRHSSLERECI, LRHSSLERECIEE, SSLERECIEEICD, ECIEEICDFEEAK, IEEICDFEEAKEI, EEICDFEEAKEIF, EICDFEEAKEIFQ, CDFEEAKEIFQNV, KEIFQNVDDTLAF, EIFQNVDDTLAFW, IFQNVDDTLAFWS, QNVDDTLAFWSKH, DDTLAFWSKHVDG, DTLAFWSKHVDGD, LAFWSKHVDGDQC, AFWSKHVDGDQCL, WSKHVDGDQCLVL, KHVDGDQCLVLPL, QCLVLPLEHPCAS, CLVLPLEHPCASL, LVLPLEHPCASLC, LPLEHPCASLCCG, ASLCCGHGTCIDG, HGTCIDGIGSFSC, TCIDGIGSFSCDC, DGIGSFSCDCRSG, GSFSCDCRSGWEG, CRSGWEGRFCQRE, SGWEGRFCQREVS, GWEGRFCQREVSF, GRFCQREVSFLNC, RFCQREVSFLNCS, QREVSFLNCSLDN, REVSFLNCSLDNG, VSFLNCSLDNGGC, SFLNCSLDNGGCT, CSLDNGGCTHYCL, THYCLEEVGWRRC, YCLEEVGWRRCSC, EEVGWRRCSCAPG, VGWRRCSCAPGYK, RRCSCAPGYKLGD, APGYKLGDDLLQC, PGYKLGDDLLQCH, YKLGDDLLQCHPA, LGDDLLQCHPAVK, GDDLLQCHPAVKF, DDLLQCHPAVKFP, DLLQCHPAVKFPC, PAVKFPCGRPWKR, VKFPCGRPWKRME, RPWKRMEKKRSHL

EXAMPLE 14 (subtilisins)

The subtilisins are a class of protease enzyme with significant economic and industrial importance. They may be used as components of detergents or cosmetics, or in the production of textiles and other industries and consumer preparations. Exposure of particular human subjects to bacterial subtilisins may evoke an unwanted hypersensitivity reaction in those individuals. There is a need for subtilisin analogues with enhanced properties and especially, improvements in the biological properties of the protein. In this regard, it is highly desired to

provide subtilisins with reduced or absent potential to induce an immune response in the human subject. Subtilisin proteins such as identified from other sources including bacterial, fungal or vertebrate sources, including mammalian organisms and man, have in common many of the peptide sequences of the present disclosure and have in common many peptide sequences with substantially the same sequence as those of the disclosed listing. Such protein sequences equally therefore fall under the scope of the present invention. Others have provided subtilisin molecules including modified subtilisins [US,5,700,676;US,4914,031; US,5,397,705; US,5,972,682].

Peptide sequences in B.*lentus* subtilisin with potential human MHC class II binding activity are:

QSVPWGISRVQAP, SVPWGISRVQAPA, WGISRVQAPAAHN, SRVQAPAAHNRGL,
VQAPAAHNRGLTG, AHNRGLTGSGVKV, RGLTGSGVKVAVL, SGVKVAVLDTGIS,
GVKVAVLDTGIST, VKVAVLDTGISTH, VAVLDTGISTHPD, AVLDTGISTHPDL,
TGISTHPDLNIRG, ISTHPDLNIRGGA, HPDLNIRGGASFV, PDLNIRGGASFVP,
LNIRGGASFVPGE, ASFVPGEPSTQDG, SFVPGEPSTQDGN, EPSTQDGNGHGTH,
GHGTHVAGTIAAL, HGTHVAGTIAALN, THVAGTIAALNNS, AGTIAALNNSIGV,
GTIAALNNSIGVL, AALNNSIGVLGVA, ALNNSIGVLGVAP, NSIGVLGVAPSAE,
GVLGVAPSAELYA, LGVAPSAELYAVK, APSAELYAVKVLG, AELYAVKVLGASG,
ELYAVKVLGASGS, YAVKVLGASGSGS, VKVLGASGSGSVS, KVLGASGSGSVSS,
SGSGSVSSIAQGL, SGSVSSIAQGLEW, GSVSSIAQGLEWA, SSIAQGLEWAGNN,
QGLEWAGNNGMHV, LEWAGNNGMHVAN, NNGMHVANLSLGS, NGMHVANLSLGSP,
MHVANLSLGSPSP, HVANLSLGSPSPS, VANLSLGSPSPSA, ANLSLGSPSPSAT,
LSLGSPSPSATLE, SPSPSATLEQAVN, SPSATLEQAVNSA, PSATLEQAVNSAT,
ATLEQAVNSATSR, TLEQAVNSATSRG, QAVNSATSRGVLV, RGVLVVAASGNSG,
GVLVVAASGNSGA, VLVVAASGNSGAG, LVVAASGNSGAGS, VAASGNSGAGSIS,
GSISYPARYANAM, ISYPARYANAMAV, YPARYANAMAVGA, ARYANAMAVGATD,
NAMAVGATDQNNN, MAVGATDQNNNRA, AVGATDQNNNRAS, NNRASFSQYGAGL,
RASFSQYGAGLDI, ASFSQYGAGLDIV, SQYGAGLDIVAPG, GAGLDIVAPGVNV,
AGLDIVAPGVNVQ, LDIVAPGVNVQST, DIVAPGVNVQSTY, APGVNVQSTYPGS,
PGVNVQSTYPGST, VNVQSTYPGSTYA, STYPGSTYASLNG, STYASLNGTSMAT,
ASLNGTSMATPHV, NGTSMATPHVAGA, MATPHVAGAAALV, TSMATPHVAGAAA,
PHVAGAAALVKQK, AALVKQKNPSWSN, ALVKQKNPSWSNV, PSWSNVQIRNHLK,
WSNVQIRNHLKNT, SNVQIRNHLKNTA, VQIRNHLKNTATS, QIRNHLKNTATSL,
RNHLKNTATSLGS, NHLKNTATSLGST, HLKNTATSLGSTN, ATSLGSTNLYGSG,
TSLGSTNLYGSGL, LGSTNLYGSGLVN, TNLYGSGLVNAEA, NLYGSGLVNAEAA,
LYGSGLVNAEAAT,

Peptide sequences in B.*amyloliquefaciens* subtilisin with potential human MHC class II binding activity are:

QSVPYGVSQIKAP, SVPYGVSQIKAPA, VPYGVSQIKAPAL, YGVSQIKAPALHS,
VSQIKAPALHSQG, SQIKAPALHSQGY, PALHSQGYTGSNV, QGYTGSNVKVAVI,
SNVKVAVIDSGID, VKVAVIDSGIDSS, KVAVIDSGIDSSH, VAVIDSGIDSSHP,
AVIDSGIDSSHPD, VIDSGIDSSHPDL, SGIDSSHPDLKVA, DSSHPDLKVAGGA,
SHPDLKVAGGASM, HPDLKVAGGASMV, PDLKVAGGASMVP, DLKVAGGASMVPS,
LKVAGGASMVPSE, GGASMVPSETNPF, ASMVPSETNPFQD, SMVPSETNPFQDN,
NPFQDNNSHGTHV, FQDNNSHGTHVAG, SHGTHVAGTVAAL, HGTHVAGTVAALN,
THVAGTVAALNNS, AGTVAALNNSIGV, GTVAALNNSIGVL, AALNNSIGVLGVA,
ALNNSIGVLGVAP, NNSIGVLGVAPSA, NSIGVLGVAPSAS, SIGVLGVAPSASL,
IGVLGVAPSASLY, GVLGVAPSASLYA, LGVAPSASLYAVK, APSASLYAVKVLG,
ASLYAVKVLGADG, SLYAVKVLGADGS, YAVKVLGADGSGQ, VKVLGADGSGQYS,
KVLGADGSGQYSW, ADGSGQYSWIING, GQYSWIINGIEWA, YSWIINGIEWAIA,
SWIINGIEWAIAN, WIINGIEWAIANN, NGIEWAIANNMDV, IEWAIANNMDVIN,
WAIANNMDVINMS, ANNMDVINMSLGG, NNMDVINMSLGGP, MDVINMSLGGPSG,
DVINMSLGGPSGS, INMSLGGPSGSAA, MSLGGPSGSAALK, AALKAAVDKAVAS,
ALKAAVDKAVASG, AAVDKAVASGVVV, AVDKAVASGVVVV, KAVASGVVVVAAA,
SGVVVVAAAGNEG, GVVVVAAAGNEGT, VVVVAAAGNEGTS, VVVAAAGNEGTSG,
AAAGNEGTSGSSS, SSTVGYPGKYPSV, STVGYPGKYPSVI, VGYPGKYPSVIAV,
GKYPSVIAVGAVD, PSVIAVGAVDSSN, SVIAVGAVDSSNQ, IAVGAVDSSNQRA,
GAVDSSNQRASFS, VDSSNQRASFSSV, ASFSSVGPELDVM, SSVGPELDVMAPG,
GPELDVMAPGVSI, PELDVMAPGVSIQ, ELDVMAPGVSIQS, LDVMAPGVSIQST,
DVMAPGVSIQSTL, APGVSIQSTLPGN, PGVSIQSTLPGNK, VSIQSTLPGNKYG,
STLPGNKYGAYNG, GNKYGAYNGTSMA, NKYGAYNGTSMAS, GAYNGTSMASPHV,
YNGTSMASPHVAG, TSMASPHVAGAAA, MASPHVAGAAALI, PHVAGAAALILSK,
AAALILSKHPNWT, AALILSKHPNWTN, ALILSKHPNWTNT, LILSKHPNWTNTQ,
PNWTNTQVRSSLE, TQVRSSLENTTTK, QVRSSLENTTTKL, VRSSLENTTTKLG,
SSLENTTTKLGDS, TKLGDSFYYGKGL, LGDSFYYGKGLIN, DSFYYGKGLINVQ,
SFYYGKGLINVQA, FYYGKGLINVQAA, YYGKGLINVQAAA

Peptide sequences in B.*subtilis* subtilisin with potential human MHC class II binding activity are:

QSVPYGISQIKAP, SVPYGISQIKAPA, VPYGISQIKAPAL, YGISQIKAPALHS,
ISQIKAPALHSQG, SQIKAPALHSQGY, PALHSQGYTGSNV, QGYTGSNVKVAVI,
SNVKVAVIDSGID, VKVAVIDSGIDSS, KVAVIDSGIDSSH, VAVIDSGIDSSHP,
AVIDSGIDSSHPD, VIDSGIDSSHPDL, SGIDSSHPDLNVR, DSSHPDLNVRGGA,
HPDLNVRGGASFV, PDLNVRGGASFVP, DLNVRGGASFVPS, LNVRGGASFVPSE,
GGASFVPSETNPY, ASFVPSETNPYQD, SFVPSETNPYQDG, NPYQDGGSHGTHV,
SHGTHVAGTIAAL, HGTHVAGTIAALN, THVAGTIAALNNS, AGTIAALNNSIGV,
GTIAALNNSIGVL, AALNNSIGVLGVS, ALNNSIGVLGVSP, NNSIGVLGVSPSA,
NSIGVLGVSPSAS, IGVLGVSPSASLY, GVLGVSPSASLYA, LGVSPSASLYAVK,
SPSASLYAVKVLD, ASLYAVKVLDSTG, YAVKVLDSTGSGQ, VKVLDSTGSGQYS,

KVLDSTGSGQYSW, STGSGQYSWIING, GQYSWIINGIEWA, YSWIINGIEWAIS,
SWIINGIEWAISN, WIINGIEWAISNN, NGIEWAISNNMDV, IEWAISNNMDVIN,
WAISNNMDVINMS, SNNMDVINMSLGG, NNMDVINMSLGGP, MDVINMSLGGPTG,
DVINMSLGGPTGS, INMSLGGPTGSTA, MSLGGPTGSTALK, TALKTVVDKAVSS,
ALKTVVDKAVSSG, KTVVDKAVSSGIV, TVVDKAVSSGIVV, VVDKAVSSGIVVA,
KAVSSGIVVAAAA, VSSGIVVAAAAGN, SGIVVAAAAGNEG, GIVVAAAAGNEGS,
IVVAAAAGNEGSS, VVAAAAGNEGSSG, AAAGNEGSSGSTS, TSTVGYPAKYPST,
STVGYPAKYPSTI, VGYPAKYPSTIAV, AKYPSTIAVGAVN, PSTIAVGAVNSSN,
STIAVGAVNSSNQ, TIAVGAVNSSNQR, IAVGAVNSSNQRA, GAVNSSNQRASFS,
VNSSNQRASFSSA, NQRASFSSAGSEL, ASFSSAGSELDVM, GSELDVMAPGVSI,
ELDVMAPGVSIQS, SELDVMAPGVSIQ, LDVMAPGVSIQST, DVMAPGVSIQSTL,
APGVSIQSTLPGG, PGVSIQSTLPGGT, VSIQSTLPGGTYG, STLPGGTYGAYNG,
GGTYGAYNGTSMA, GTYGAYNGTSMAT, GAYNGTSMATPHV, YNGTSMATPHVAG,
TSMATPHVAGAAA, MATPHVAGAAALI, PHVAGAAALILSK, GAAALILSKHPTW,
AALILSKHPTWTN, ALILSKHPTWTNA, LILSKHPTWTNAQ, PTWTNAQVRDRLE,
AQVRDRLESTATY, QVRDRLESTATYL, DRLESTATYLGNS, ATYLGNSFYYGKG,
TYLGNSFYYGKGL, LGNSFYYGKGLIN, NSFYYGKGLINVQ, SFYYGKGLINVQA,
FYYGKGLINVQAA, YYGKGLINVQAAA

Peptide sequences in B. *licheniformis* subtilisin with potential human MHC class II binding activity are:

QTVPYGIPLIKAD, VPYGIPLIKADKV, YGIPLIKADKVQA, IPLIKADKVQAQG,
PLIKADKVQAQGF, IKADKVQAQGFKG, DKVQAQGFKGANV, QGFKGANVKVAVL,
ANVKVAVLDTGIQ, VKVAVLDTGIQAS, KVAVLDTGIQASH, VAVLDTGIQASHP,
AVLDTGIQASHPD, VLDTGIQASHPDL, DTGIQASHPDLNV, TGIQASHPDLNVV,
QASHPDLNVVGGA, HPDLNVVGGASFV, PDLNVVGGASFVA, DLNVVGGASFVAG,
LNVVGGASFVAGE, NVVGGASFVAGEA, ASFVAGEAYNTDG, SFVAGEAYNTDGN,
EAYNTDGNGHGTH, GHGTHVAGTVAAL, HGTHVAGTVAALD, THVAGTVAALDNT,
GTVAALDNTTGVL, TVAALDNTTGVLG, AALDNTTGVLGVA, DNTTGVLGVAPSV,
TTGVLGVAPSVSL, TGVLGVAPSVSLY, GVLGVAPSVSLYA, LGVAPSVSLYAVK,
APSVSLYAVKVLN, PSVSLYAVKVLNS, VSLYAVKVLNSSG, SLYAVKVLNSSGS,
YAVKVLNSSGSGS, VKVLNSSGSGSYS, KVLNSSGSGSYSG, GSYSGIVSGIEWA,
TNGMDVINMSLGG, NGMDVINMSLGGA, MDVINMSLGGASG, DVINMSLGGASGS,
INMSLGGASGSTA, MSLGGASGSTAMK, TAMKQAVDNAYAR, AMKQAVDNAYARG,
QAVDNAYARGVVV, NAYARGVVVVAAA, RGVVVVAAAGNSG, GVVVVAAAGNSGN,
VVVVAAAGNSGNS, VVVAAAGNSGNSG, NTIGYPAKYDSVI, IGYPAKYDSVIAV,
AKYDSVIAVGAVD, DSVIAVGAVDSNS, SVIAVGAVDSNSN, IAVGAVDSNSNRA,
AVGAVDSNSNRAS, GAVDSNSNRASFS, AVDSNSNRASFSS, SNRASFSSVGAEL,
ASFSSVGAELEVM, SSVGAELEVMAPG, GAELEVMAPGAGV, AELEVMAPGAGVY,
ELEVMAPGAGVYS, LEVMAPGAGVYST, EVMAPGAGVYSTY, APGAGVYSTYPTN,
AGVYSTYPTNTYA, GVYSTYPTNTYAT, STYPTNTYATLNG, NTYATLNGTSMAS,

ATLNGTSMASPHV, LNGTSMASPHVAG, TSMASPHVAGAAA, MASPHVAGAAALI,
PHVAGAAALILSK, GAAALILSKHPNL, AALILSKHPNLSA, ALILSKHPNLSAS,
LILSKHPNLSASQ, SKHPNLSASQVRN, HPNLSASQVRNRL, PNLSASQVRNRLS,
LSASQVRNRLSST, SQVRNRLSSTATY, QVRNRLSSTATYL, NRLSSTATYLGSS,
ATYLGSSFYYGKG, TYLGSSFYYGKGL, LGSSFYYGKGLIN, SSFYYGKGLINVE,
SFYYGKGLINVEA, FYYGKGLINVEAA, YYGKGLINVEAAA

EXMAMPLE 15 (ligands of CNTF):

The present invention provides for modified forms of the protein subunits comprising a heterodimeric ligand for the ciliary neurotrophic factor (CNTF) receptor complex in humans. The receptor complex is activated by at least two ligands including CNTF and a heterodimeric complex comprising cardiotrophin-like cytokine (CLC) and the soluble receptor cytokine-like factor 1 (CLF) [Elson G. C. A. et al (2000) *Nature Neuroscience* 3: 867-872]. CLC is a protein of the IL-6 family of cytokines and is also known as novel neurotrophin-1/B cell-stimulating factor-3 [Senaldi, G. et al (1999) *Proc. Nat. Acad. Sci. USA* 96: 11458-11463, US,5,741,772]. CLF is homologous to proteins of the cytokine type I receptor family [Elson, G. C. A. et al (1998) *Journal of Immunol.* 161: 1371-1379] and has also been identified as NR6 [Alexander W.S. et al (1999) *Curr. Biol.* 9: 605-608]. Heterodimers formed by association of CLC and CLF have been shown to directly interact with the CNTFR and the so formed trimeric complex is able to stimulate signalling events within cells expressing the other recognised components of the CNTFR complex such as gp130 and LIFR [Elson G. C. A. et al (2000) *ibid*].

Peptide sequences in human CLC with potential human MHC class II binding activity are:

PGPSIQKTYDLTR, PSIQKTYDLTRYL, IQKTYDLTRYLEH, KTYDLTRYLEHQL,
YDLTRYLEHQLRS, LTRYLEHQLRSLA, TRYLEHQLRSLAG, RYLEHQLRSLAGT,
HQLRSLAGTYLNY, QLRSLAGTYLNYL, RSLAGTYLNYLGP, GTYLNYLGPPFNE,
TYLNYLGPPFNEP, NYLGPPFNEPDFN, PPFNEPDFNPPRL, PFNEPDFNPPRLG,
PDFNPPRLGAETL, FNPPRLGAETLPR, PRLGAETLPRATV, LGAETLPRATVDL,
ETLPRATVDLEVW, PRATVDLEVWRSL, ATVDLEVWRSLND, TVDLEVWRSLNDK,
VDLEVWRSLNDKL, LEVWRSLNDKLRL, EVWRSLNDKLRLT, VWRSLNDKLRLTQ,
RSLNDKLRLTQNY, DKLRLTQNYEAYS, KLRLTQNYEAYSH, LRLTQNYEAYSHL,
TQNYEAYSHLLCY, QNYEAYSHLLCYL, EAYSHLLCYLRGL, SHLLCYLRGLNRQ,
HLLCYLRGLNRQA, LCYLRGLNRQAAT, CYLRGLNRQAATA, RGLNRQAATAELR,
GLNRQAATAELRR, QAATAELRRSLAH, AATAELRRSLAHF, AELRRSLAHFCTS,
ELRRSLAHFCTSL, RSLAHFCTSLQGL, AHFCTSLQGLLGS, TSLQGLLGSIAGV,
SLQGLLGSIAGVM, QGLLGSIAGVMAA, GLLGSIAGVMAAL, LLGSIAGVMAALG,
GSIAGVMAALGYP, SIAGVMAALGYPL, AGVMAALGYPLPQ, GVMAALGYPLPQP,

AALGYPLPQPLPG, LGYPLPQPLPGTE, YPLPQPLPGTEPT, QPLPGTEPTWTPG,
PTWTPGPAHSDFL, WTPGPAHSDFLQK, HSDFLQKMDDFWL, SDFLQKMDDFWLL,
DFLQKMDDFWLLK, FLQKMDDFWLLKE, QKMDDFWLLKELQ, DDFWLLKELQTWL,
DFWLLKELQTWLW, FWLLKELQTWLWR, WLLKELQTWLWRS, KELQTWLWRSAKD,
ELQTWLWRSAKDF, QTWLWRSAKDFNR, TWLWRSAKDFNRL, WLWRSAKDFNRLK,
WRSAKDFNRLKKK, RSAKDFNRLKKKM, KDFNRLKKKMQPP, NRLKKKMQPPAAA,
RLKKKMQPPAAAV, LKKKMQPPAAAVT, KKMQPPAAAVTLH, KMQPPAAAVTLHL,
QPPAAAVTLHLGA

Peptide sequences in human CLF with potential human MHC class II binding activity are:

TAVISPQDPTLLI, AVISPQDPTLLIG, VISPQDPTLLIGS, QDPTLLIGSSLLA,
DPTLLIGSSLLAT, PTLLIGSSLLATC, TLLIGSSLLATCS, LLIGSSLLATCSV,
IGSSLLATCSVHG, SSLLATCSVHGDP, SLLATCSVHGDPP, CSVHGDPPGATAE,
GDPPGATAEGLYW, EGLYWTLNGRRLP, GLYWTLNGRRLPP, WTLNGRRLPPELS,
RRLPPELSRVLNA, RLPPELSRVLNAS, PELSRVLNASTLA, ELSRVLNASTLAL,
LSRVLNASTLALA, SRVLNASTLALAL, RVLNASTLALALA, LNASTLALALANL,
NASTLALALANLN, STLALALANLNGS, LALALANLNGSRQ, LALANLNGSRQRS,
ANLNGSRQRSGDN, DNLVCHARDGSIL, NLVCHARDGSILA, VCHARDGSILAGS,
RDGSILAGSCLYV, DGSILAGSCLYVG, GSILAGSCLYVGL, SILAGSCLYVGLP,
SCLYVGLPPEKPV, CLYVGLPPEKPVN, LYVGLPPEKPVNI, VGLPPEKPVNISC,
KPVNISCWSKNMK, VNISCWSKNMKDL, KNMKDLTCRWTPG, KDLTCRWTPGAHG,
CRWTPGAHGETFL, RWTPGAHGETFLH, HGETFLHTNYSLK, ETFLHTNYSLKYK,
TFLHTNYSLKYKL, TNYSLKYKLRWYG, YSLKYKLRWYGQD, LKYKLRWYGQDNT,
YKLRWYGQDNTCE, LRWYGQDNTCEEY, RWYGQDNTCEEYH, EEYHTVGPHSCHI,
HTVGPHSCHIPKD, PHSCHIPKDLALF, CHIPKDLALFTPY, IPKDLALFTPYEI,
KDLALFTPYEIWV, ALFTPYEIWVEAT, TPYEIWVEATNRL, YEIWVEATNRLGS,
EIWVEATNRLGSA, IWVEATNRLGSAR, EIWVEATNRLGSA, NRLGSARSDVLTL,
EATNRLGSARSDV, SARSDVLTLDILD, SDVLTLDILDVVT, DVLTLDILDVVTT,
LTLDILDVVTTDP, LDILDVVTTDPPP, DILDVVTTDPPPD, LDVVTTDPPPDVH,
ARSDVLTLDILDV, PDVHVSRVGGLED, VHVSRVGGLEDQL, SRVGGLEDQLSVR,
DVVTTDPPPDVHV, GGLEDQLSVRWVS, RVGGLEDQLSVRW, DQLSVRWVSPPAL,
LSVRWVSPPALKD, VRWVSPPALKDFL, RWVSPPALKDFLF, GLEDQLSVRWVSP,
PALKDFLFQAKYQ, KDFLFQAKYQIRY, DFLFQAKYQIRYR, FLFQAKYQIRYRV,
VSPPALKDFLFQA, AKYQIRYRVEDSV, YQIRYRVEDSVDW, IRYRVEDSVDWKV,
YRVEDSVDWKVVD, FQAKYQIRYRVED, DSVDWKVVDDVSN, VDWKVVDDVSNQT,
VEDSVDWKVVDDV, KVVDDVSNQTSCR, DDVSNQTSCRLAG, WKVVDDVSNQTSC,
QTSCRLAGLKPGT, CRLAGLKPGTVYF, AGLKPGTVYFVQV, GTVYFVQVRCNPF,
TVYFVQVRCNPFG, VYFVQVRCNPFGI, FVQVRCNPFGIYG, VQVRCNPFGIYGS,
NPFGIYGSKKAGI, PFGIYGSKKAGIW, FGIYGSKKAGIWS, GIYGSKKAGIWSE,
SKKAGIWSEWSHP, AGIWSEWSHPTAA, GIWSEWSHPTAAS, SEWSHPTAASTPR,
SHPTAASTPRSER, PSSGPVRRELKQF, GPVRRELKQFLGW, RELKQFLGWLKKH,

KQFLGWLKKHAYC, QFLGWLKKHAYCS, LGWLKKHAYCSNL, GWLKKHAYCSNLS,
HAYCSNLSFRLYD, AYCSNLSFRLYDQ, SNLSFRLYDQWRA, LSFRLYDQWRAWM,
FRLYDQWRAWMQK, RLYDQWRAWMQKS, DQWRAWMQKSHKT, RAWMQKSHKTRNQ,
AWMQKSHKTRNQD, HKTRNQDEGILPS, EGILPSGRRGTAR, GILPSGRRGTARG,

EXAMPLE 16 (follicle-stimulating hormone)

The present invention provides for modified forms of human hFSH with one or more T cell epitopes removed. hFSH is a glycoprotein hormone with a dimeric structure containing two glycoprotein subunits. The protein is being used therapeutically in the treatment of human infertility and a recombinant form of the protein has been the subject of a number of clinical trials [Out, H.J. et al (1995) *Hum. Reprod.* 10: 2534-2540; Hedon, B. et al (1995) *Hum. Reprod.* 10: 3102-3106; Recombinant Human FSH study Group (1995) *Fertil. Steril.* 63:77-86; Prevost, R.R. (1998) *Pharmacotherapy* 18: 1001-1010].

Peptide sequences in human hFSH with potential human MHC class II binding activity are:

KTLQFFFLFCCWK, LQFFFLFCCWKAI, QFFFLFCCWKAIC, FFFLFCCWKAICC,
FFLFCCWKAICCN, FLFCCWKAICCNS, CCWKAICCNSCEL, KAICCNSCELTNI,
CELTNITIAIEKE, TNITIAIEKEECR, ITIAIEKEECRFC, IAIEKEECRFCIS,
CRFCISINTTWCA, FCISINTTWCAGY, ISINTTWCAGYCY, TTWCAGYCYTRDL,
AGYCYTRDLVYKD, YCYTRDLVYKDPA, RDLVYKDPARPKI, DLVYKDPARPKIQ,
LVYKDPARPKIQK, PKIQKTCTFKELV, CTFKELVYETVRV, KELVYETVRVPGC,
ELVYETVRVPGCA, LVYETVRVPGCAH, ETVRVPGCAHHAD, VRVPGCAHHADSL,
DSLYTYPVATQCH, SLYTYPVATQCHC, YTYPVATQCHCGK, YPVATQCHCGKCD,
CTVRGLGPSYCSF, RGLGPSYCSFGEM

EXAMPLE 16 (ricin A)

The present invention provides for modified forms of ricin toxin A-chain (RTA) with one or more T cell epitopes removed. Ricin is a cytotoxin originally isolated from the seeds of the castor plant and is an example of a type II ribosome inactivating protein (RIP). The native mature protein is a heterodimer comprising the RTA of 267 amino acid residues in disulphide linkage with the ricin B-chain of 262 amino acid residues. The B-chain is a lectin with binding affinity for galactosides. The native protein is able to bind cells via the B-chain and enters the cell by endocytosis. Inside the cell, the RTA is released from the B-chain by reduction of the disulphide linkage and is released from the endosome into the cytoplasm via unknown mechanisms. In the cytoplasm the toxin degrades ribosomes by action as a specific N-glycosylase rapidly resulting in the cessation of protein translation and cell death. The

extreme cytotoxicity of RTA and other RIPs has lead to their use in experimental therapies for the treatment of cancer and other diseases where ablation of a particular cell population is required. Immunotoxin molecules containing antibody molecules in linkage with RTA have been produced and used in a number of clinical trials [Ghetie, M.A. et al (1991) *Cancer Res.* 51: 5876-5880; Vitetta, E.S. et al (1991) *Cancer Res.* 51: 4052-4058; Amlot, P.L. et al (1993) *Blood* 82: 2624-2633; Conry, R.M. et al (1995) *J. Immunother. Emphasis Tumor Immunol.* 18: 231-241; Schnell, R. et al (2000) *Leukaemia* 14: 129-135]. In the immunotoxin the antibody domain provides binding to the surface of the desired target cell and linkage to the RTA may be via chemical cross-linkage or as a recombinant fusion protein.

Peptide sequences in ricin toxin a-chain with potential human MHC class II binding activity are:

KQYPIINFTTAGA, YPIINFTTAGATV, PIINFTTAGATVQ, INFTTAGATVQSY,
ATVQSYTNFIRAV, QSYTNFIRAVRGR, TNFIRAVRGRLTT, NFIRAVRGRLTTG,
RAVRGRLTTGADV, GRLTTGADVRHEI, ADVRHEIPVLPNR, HEIPVLPNRVGLP,
IPVLPNRVGLPIN, PVLPNRVGLPINQ, NRVGLPINQRFIL, VGLPINQRFILVE,
LPINQRFILVELS, QRFILVELSNHAE, RFILVELSNHAEL, FILVELSNHAELS,
ILVELSNHAELSV, VELSNHAELSVTL, AELSVTLALDVTN, LSVTLALDVTNAY,
VTLALDVTNAYVV, LALDVTNAYVVGY, LDVTNAYVVGYRA, NAYVVGYRAGNSA,
AYVVGYRAGNSAY, YVVGYRAGNSAYF, VGYRAGNSAYFFH, SAYFFHPDNQEDA,
AYFFHPDNQEDAE, YFFHPDNQEDAEA, EAITHLFTDVQNR, THLFTDVQNRYTF,
HLFTDVQNRYTFA, TDVQNRYTFAFGG, NRYTFAFGGNYDR, YTFAFGGNYDRLE,
FAFGGNYDRLEQL, GNYDRLEQLAGNL, DRLEQLAGNLREN, EQLAGNLRENIEL,
GNLRENIELGNGP, ENIELGNGPLEEA, IELGNGPLEEAIS, GPLEEAISALYYY,
EAISALYYYSTGG, SALYYYSTGGTQL, ALYYYSTGGTQLP, LYYYSTGGTQLPT,
YYYSTGGTQLPTL, TQLPTLARSFIIC, PTLARSFIICIQM, RSFIICIQMISEA,
SFIICIQMISEAA, FIICIQMISEAAR, ICIQMISEAARFQ, IQMISEAARFQYI,
QMISEAARFQYIE, ARFQYIEGEMRTR, FQYIEGEMRTRIR, QYIEGEMRTRIRY,
GEMRTRIRYNRRS, TRIRYNRRSAPDP, IRYNRRSAPDPSV, PSVITLENSWGRL,
SVITLENSWGRLS, ITLENSWGRLSTA, NSWGRLSTAIQES, GRLSTAIQESNQG,
TAIQESNQGAFAS, GAFASPIQLQRRN, SPIQLQRRNGSKF, IQLQRRNGSKFSV,
SKFSVYDVSILIP, FSVYDVSILIPII, SVYDVSILIPIIA, YDVSILIPIIALM,
VSILIPIIALMVY, SILIPIIALMVYR, IPIIALMVYRCAP, IALMVYRCAPPPS,
ALMVYRCAPPPSS, LMVYRCAPPPSSQ, MVYRCAPPPSSQF

EXAMPLE 17 (adipocyte complement-related protein):

The present invention provides for modified forms of human or mouse Acrp30 with one or more T cell epitopes removed. Acrp30 is an abundant serum protein of approximately 30kDa

molecular weight expressed exclusively by adipocyte cells [Scherer, P. E. et al (1995) *J. Biol. Chem.* 270: 26746-26749]. The human gene Acrp30 protein sequence is disclosed e.g. in US, 5,869,330. Secretion of the protein is enhanced by insulin and levels of the protein are decreased in obese subjects. The protein is involved in the regulation of energy balance and in particular the regulation of fatty acid metabolism. Four sequence domains are identified in the mouse and human protein comprising a cleaved N-terminal signal, a region with no recognized homology to other proteins, a collagen-like domain and a globular domain. The globular domain may be removed from the mouse protein by protease treatment to produce gAcrp30. Preparations of murine gAcrp30 have pharmaceutical properties and have been shown to decrease elevated levels of free fatty acids in the serum of mice following administration of high fat meals or i.v. injection of lipid [Fruebis, J. et al (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98: 2005-2010].

Peptide sequences in mouse Acrp30 with potential human MHC class II binding activity are:.

DDVTTTEELAPAL, TTTEELAPALVPP, EELAPALVPPPKG, LAPALVPPPKGTC,
PALVPPPKGTCAG, ALVPPPKGTCAGW, AGWMAGIPGHPGH, GWMAGIPGHPGHN,
AGIPGHPGHNGTP, GTPGRDGRDGTPG, GDAGLLGPKGETG, AGLLGPKGETGDV,
GLLGPKGETGDVG, GETGDVGMTGAEG, GDVGMTGAEGPRG, VGMTGAEGPRGFP,
RGFPGTPGRKGEP, TPGRKGEPGEAAY, GRKGEPGEAAYMY, AAYMYRSAFSVGL,
AYMYRSAFSVGLE, YMYRSAFSVGLET, SAFSVGLETRVTV, FSVGLETRVTVPN,
VGLETRVTVPNVP, GLETRVTVPNVPI, ETRVTVPNVPIRF, TRVTVPNVPIRFT,
VTVPNVPIRFTKI, VPNVPIRFTKIFY, PNVPIRFTKIFYN, VPIRFTKIFYNQQ,
IRFTKIFYNQQNH, RFTKIFYNQQNHY, TKIFYNQQNHYDG, KIFYNQQNHYDGS,
IFYNQQNHYDGST, QQNHYDGSTGKFY, NHYDGSTGKFYCN, GKFYCNIPGLYYF,
KFYCNIPGLYYFS, CNIPGLYYFSYHI, PGLYYFSYHITVY, GLYYFSYHITVYM,
LYYFSYHITVYMK, YYFSYHITVYMKD, FSYHITVYMKDVK, SYHITVYMKDVKV,
YHITVYMKDVKVS, HITVYMKDVKVSL, ITVYMKDVKVSLF, TVYMKDVKVSLFK,
VYMKDVKVSLFKK, KDVKVSLFKKDKA, VKVSLFKKDKAVL, VSLFKKDKAVLFT,
SLFKKDKAVLFTY, FKKDKAVLFTYDQ, KDKAVLFTYDQYQ, KAVLFTYDQYQEK,
AVLFTYDQYQEKN, VLFTYDQYQEKNV, FTYDQYQEKNVDQ, YDQYQEKNVDQAS,
DQYQEKNVDQASG, EKNVDQASGSVLL, KNVDQASGSVLLH, ASGSVLLHLEVGD,
GSVLLHLEVGDQV, SVLLHLEVGDQVW, VLLHLEVGDQVWL, LHLEVGDQVWLQV,
LEVGDQVWLQVYG, DQVWLQVYGDGDH, QVWLQVYGDGDHN, VWLQVYGDGDHNG,
LQVYGDGDHNGLY, QVYGDGDHNGLYA, VYGDGDHNGLYAD, GDHNGLYADNVND,
NGLYADNVNDSTF, GLYADNVNDSTFT, LYADNVNDSTFTG, DNVNDSTFTGFLL,
VNDSTFTGFLLYH, STFTGFLLYHDTN

Peptide sequences in human Acrp30 with potential human MHC class II binding activity are:

PGVLLPLPKGACT, GVLLPLPKGACTG, VLLPLPKGACTGW, LPLPKGACTGWMA,
PLPKGACTGWMAG, TGWMAGIPGHPGH, GWMAGIPGHPGHN, AGIPGHPGHNGAP,
GAPGRDGRDGTPG, GDPGLIGPKGDIG, PGLIGPKGDIGET, GLIGPKGDIGETG,
GPKGDIGETGVPG, GDIGETGVPGAEG, TGVPGAEGPRGFP, RGFPGIQGRKGEP,
PGIQGRKGEPGEG, GRKGEPGEGAYVY, GAYVYRSAFSVGL, AYVYRSAFSVGLE,
YVYRSAFSVGLET, RSAFSVGLETYVT, SAFSVGLETYVTI, AFSVGLETYVTIP,
FSVGLETYVTIPN, VGLETYVTIPNMP, GLETYVTIPNMPI, ETYVTIPNMPIRF,
TYVTIPNMPIRFT, VTIPNMPIRFTKI, IPNMPIRFTKIFY, PNMPIRFTKIFYN,
MPIRFTKIFYNQQ, IRFTKIFYNQQNH, RFTKIFYNQQNHY, TKIFYNQQNHYDG,
KIFYNQQNHYDGS, IFYNQQNHYDGST, QQNHYDGSTGKFH, NHYDGSTGKFHCN,
GKFHCNIPGLYYF, CNIPGLYYFAYHI, PGLYYFAYHITVY, GLYYFAYHITVYM,
LYYFAYHITVYMK, YYFAYHITVYMKD, FAYHITVYMKDVK, AYHITVYMKDVKV,
YHITVYMKDVKVS, HITVYMKDVKVSL, ITVYMKDVKVSLF, TVYMKDVKVSLFK,
VYMKDVKVSLFKK, KDVKVSLFKKDKA, VKVSLFKKDKAML, VSLFKKDKAMLFT,
SLFKKDKAMLFTY, FKKDKAMLFTYDQ, KDKAMLFTYDQYQ, KAMLFTYDQYQEN,
AMLFTYDQYQENN, MLFTYDQYQENNV, FTYDQYQENNVDQ, YDQYQENNVDQAS,
DQYQENNVDQASG, ENNVDQASGSVLL, NNVDQASGSVLLH, ASGSVLLHLEVGD,
GSVLLHLEVGDQV, SVLLHLEVGDQVW, VLLHLEVGDQVWL, LHLEVGDQVWLQV,
LEVGDQVWLQVYG, DQVWLQVYGEGER, QVWLQVYGEGERN, VWLQVYGEGERNG,
LQVYGEGERNGLY, QVYGEGERNGLYA, NGLYADNDNDSTF, GLYADNDNDSTFT,
LYADNDNDSTFTG, DNDSTFTGFLLYH, STFTGFLLYHDTN.

EXAMPLE 18 (anti-C5 antibody):

The present invention provides for modified forms of monoclonal antibodies with binding specificity directed to the human C5 complement protein. The invention provides for modified antibodies with one or more T cell epitopes removed. The antibodies with binding specificity to C5 complement protein block cleavage activation of the C5 convertase and thereby inhibit the production of the pro-inflammatory components C5a and C5b-9. Activation of the complement system is a significant contributory factor in the pathogenesis of a number of acute and chronic diseases, and inhibition of the complement cascade at the level of C5 offers significant promise as a therapeutic avenue for some of these [Morgan B.P. (1994) *Eur. J. Clin. Invest.* 24: 219-228]. A number of anti-C5 antibodies and methods for their therapeutic use have been described in the art [Wurzner R. et al (1991) *Complement Inflamm.* 8: 328-340; Thomas, T.C. et al (1996) *Molecular Immunology* 33: 1389-14012; US,5,853,722; US, 6,074,64]. The antibody designated 5G1.1 [Thomas, T.C. et al (1996) *ibid*] and a single-chain humanised variant are undergoing clinical trials for a number of disease indications including cardiopulmonary bypass [Fitch, J.C.K. et al (1999) *Circulation* 100:

2499-2506] and rheumatoid arthritis. The invention discloses sequences identified within the anti-C5 antibody designated 5G1.1 [Thomas, T.C. et al].The sequences disclosed are derived from the variable region domains of both the heavy and light chains of the antibody sequence that are potential T cell epitopes by virtue of MHC class II binding potential. The disclosure further identifies potential epitopes within the protein sequence of a single-chain and "humanised" variant 5G1.1 antibody [Thomas, T.C. et al (1996) *ibid*].

Peptide sequences in the heavy-chain variable region of antibody 5G1.1 with potential human MHC class II binding activity are:

VQLQQSGAELMKP, QSGAELMKPGASV, AELMKPGASVKMS, ELMKPGASVKMSC, ASVKMSCKATGYI,
VKMSCKATGYIFS, KMSCKATGYIFSN, ATGYIFSNYWIQW, TGYIFSNYWIQWI, GYIFSNYWIQWIK,
YIFSNYWIQWIKQ, SNYWIQWIKQRPG, NYWIQWIKQRPGH, YWIQWIKQRPGHG, IQWIKQRPGHGLE,
QWIKQRPGHGLEW, HGLEWIGEILPGS, LEWIGEILPGSGS, EWIGEILPGSGST, WIGEILPGSGSTE,
GEILPGSGSTEYT, EILPGSGSTEYTE, TEYTENFKDKAAF, ENFKDKAAFTADT, FKDKAAFTADTSS,
KAAFTADTSSNTA, AAFTADTSSNTAY, TAYMQLSSLTSED, AYMQLSSLTSEDS, MQLSSLTSEDSAV,
SSLTSEDSAVYYC, SLTSEDSAVYYCA, TSEDSAVYYCARY, SAVYYCARYFFGS, AVYYCARYFFGSS,
VYYCARYFFGSSP, CARYFFGSSPNWY, ARYFFGSSPNWYF, RYFFGSSPNWYFD, YFFGSSPNWYFDV,
PNWYFDVWGAGTT, NWYFDVWGAGTTV, WYFDVWGAGTTVT, FDVWGAGTTVTVS, DVWGAGTTVTVSS

Peptide sequences in the light-chain variable region of antibody 5G1.1 with potential human MHC class II binding activity are:

IQMTQSPASLSAS, ASLSASVGETVTI, ASVGETVTITCGA, ETVTITCGASENI, VTITCGASENIYG,
TITCGASENIYGA, ENIYGALNWYQRK, NIYGALNWYQRKQ, GALNWYQRKQGKS, LNWYQRKQGKSPQ,
NWYQRKQGKSPQL, GKSPQLLIYGATN, PQLLIYGATNLAD, QLLIYGATNLADG, LLIYGATNLADGM,
LIYGATNLADGMS, TNLADGMSSRFSG, DGMSSRFSGSGSG, SRFSGSGSGRQYY, SGSGRQYYLKISS,
RQYYLKISSLHPD, QYYLKISSLHPDD, YYLKISSLHPDDV, LKISSLHPDDVAT, SSLHPDDVATYYC,
SLHPDDVATYYCQ, DDVATYYCQNVLN, ATYYCQNVLNTPL, TYYCQNVLNTPLT, YYCQNVLNTPLTF,
YCQNVLNTPLTFG, CQNVLNTPLTFGA, QNVLNTPLTFGAG, NVLNTPLTFGAGT, TPLTFGAGTKLEL

EXAMPLE 19 (anti-CD20 antibodies):

The present invention provides for modified forms of a monoclonal antibody with binding specificity to the human CD20 antigen. CD20 is a B-cell specific surface molecule expressed on pre-B and mature B-cells including greater than 90% of B-cell non-Hodgkin's lymphomas (NHL). Monoclonal antibodies and radioimmunoconjugates targeting of CD20 have emerged as new treatments for NHL. Significant examples include the monoclonal antibodies 2B8 [Reff, M.E. et al (1994) *Blood* 83: 435-445] and B1 [US,6,090,365]. The variable region domains of 2B8 have been cloned and combined with human constant region domains to produce a chimeric antibody designated C2B8 which is marketed as Rituxan™ in the USA

[US, 5,776,456] or MabThera^{R} (rituximab) in Europe. C2B8 is recognized as a valuable therapeutic agent for the treatment of NHL and other B-cell diseases [Maloney, D.G. et al (1997) *J. Clin. Oncol.* 15: 3266-3274; Maloney, D.G. et al (1997) *Blood* 90: 2188-2195]. The B1 antibody has similarly achieved registration for use as a NHL therapeutic although in this case the molecule (Bexxar^{TM}) is a ^{131}I radioimmunoconjugate although the native (non-conjugated) antibody has utility in *ex vivo* purging regimens for autologous bone marrow transplantation therapies for lymphoma and refractory leukemia [Freedman, A.S. et al (1990), *J. Clin. Oncol.*8: 784]. Despite the success of antibodies such as C2B8 (rituximab) and Bexxar^{TM} there is a continued need for anti-CD20 analogues with enhanced properties.

Peptide sequences in the heavy-chain variable region of antibody 2B8 with potential human MHC class II binding activity are:

VQLQQPGAELVKA, LQQPGAELVKAGA, AELVKAGASVKMS, ELVKAGASVKMSC,
ASVKMSCKASGYT, VKMSCKASGYTFT, KMSCKASGYTFTS, ASGYTFTSYNMHW,
SGYTFTSYNMHWV, YTFTSYNMHWVKQ, TSYNMHWVKQTPG, YNMHWVKQTPGRG,
MHWVKQTPGRGLE, HWVKQTPGRGLEW, TPGRGLEWIGAIY, RGLEWIGAIYPGN,
GLEWIGAIYPGNG, EWIGAIYPGNGDT, GAIYPGNGDTSYN, AIYPGNGDTSYNQ,
YPGNGDTSYNQKF, TSYNQKFKGKATL, YNQKFKGKATLTA, QKFKGKATLTADK,
ATLTADKSSSTAY, TAYMQLSSLTSED, AYMQLSSLTSEDS, MQLSSLTSEDSAV,
SSLTSEDSAVYYC, SLTSEDSAVYYCA, TSEDSAVYYCARS, SAVYYCARSTYYG,
AVYYCARSTYYGG, VYYCARSTYYGGD, STYYGGDTYFNVW, TYYGGDTYFNVWG,
DTYFNVWGAGTTV, TYFNVWGAGTTVT, FNVWGAGTTVTVS, NVWGAGTTVTVSA

Peptide sequences in the light-chain variable region of antibody 2B8 with potential human MHC class II binding activity are:

QIVLSQSPAILSA, IVLSQSPAILSAS, QSPAILSASPGEK, PAILSASPGEKVT,
AILSASPGEKVTM, EKVTMTCRASSSV, VTMTCRASSSVSY, TMTCRASSSVSYI,
SSVSYIHWFQQKP, VSYIHWFQQKPGS, SYIHWFQQKPGSS, IHWFQQKPGSSPK,
KPWIYATSNLASG, PWIYATSNLASGV, WIYATSNLASGVP, ATSNLASGVPVRF,
SNLASGVPVRFSG, SGVPVRFSGSGSG, VPVRFSGSGSGTS, VRFSGSGSGTSYS,
GTSYSLTISRVEA, TSYSLTISRVEAE, SYSLTISRVEAED, YSLTISRVEAEDA,
LTISRVEAEDAAT, SRVEAEDAATYYC, RVEAEDAATYYCQ, ATYYCQQWTSNPP,
TYYCQQWTSNPPT, QQWTSNPPTFGGG, NPPTFGGGTKLEI

EXAMPLE 20:

The present invention provides for modified forms of a monoclonal antibody with binding specificity to the human IL-2 receptor. The monoclonal antibody is designated anti-Tac and the modified form has one or more T cell epitopes removed. The anti-Tac antibody binds with

high specificity to the alpha subunit (p55-alpha, CD25 or Tac) of the human high affinity IL-2 receptor expressed on the surface of T and B lymphocytes. Antibody binding blocks the ability of IL-2 to bind the receptor and achieve T-cell activation. The ability of the anti-Tac antibody to act as an IL-2 antagonist has significant clinical potential in the treatment of organ transplant rejection. Clinical studies using the mouse antibody have shown some initial benefit to patients who have undergone kidney transplant although the long term benefit over conventional immune suppression was not found due the development of a HAMA response in a high proportion of patients [Kirkham, R.L. et al (1991) *Transplantation* 51: 107-113]. A "humanized" anti-Tac antibody has been developed in which significant components of the protein have been engineered to contain protein sequence identified from a human antibody gene [Queen, C. et al (1989) *Proc. Natl. Acad. Sci. (USA)* 86: 10029-10033; US,5,530,101; US,5,585,089; US,6,013,256]. The "humanised" anti-Tac (Zenapax™ or daclizumab) has undergone clinical trials as an immune suppressive agent for the management of acute graft versus host disease and suppression of kidney transplant rejection [Anasetti, C. et al (1994), Blood 84: 1320-1327; Anasetti, C. et al (1995) *Blood* 86: Supplement 1:62a; Eckhoff, D.E. et al (2000) *Transplantation* 69: 1867-1872; Ekberg, H. et al (1999) *Transplant Proc.* 31: 267-268].

Peptide sequences in the heavy-chain variable region of mouse anti-Tac antibody with potential human MHC class II binding activity are:

VQLQQSGAELAKP, AELAKPGASVKMS, ASVKMSCKASGYT, VKMSCKASGYTFT,
KMSCKASGYTFTS, ASGYTFTSYRMHW, SGYTFTSYRMHWV, YTFTSYRMHWVKQ,
TSYRMHWVKQRPG, YRMHWVKQRPGQG, MHWVKQRPGQGLE, HWVKQRPGQGLEW,
RPGQGLEWIGYIN, QGLEWIGYINPST, LEWIGYINPSTGY, EWIGYINPSTGYT,
IGYINPSTGYTEY, GYINPSTGYTEYN, TGYTEYNQKFKDK, TEYNQKFKDKATL,
QKFKDKATLTADK, ATLTADKSSSTAY, TAYMQLSSLTFED, AYMQLSSLTFEDS,
YMQLSSLTFEDSA, MQLSSLTFEDSAV, SSLTFEDSAVYYC, SLTFEDSAVYYCA,
LTFEDSAVYYCAR, SAVYYCARGGGVF, AVYYCARGGGVFD, VYYCARGGGVFDY,
GGVFDYWGQGTTL, GVFDYWGQGTTLT, FDYWGQGTTLTVS, DYWGQGTTLTVSS

Peptide sequences in the light-chain variable region of mouse anti-Tac antibody with potential human MHC class II binding activity are:

QIVLTQSPAIMSA, IVLTQSPAIMSAS, QSPAIMSASPGEK, PAIMSASPGEKVT,
AIMSASPGEKVTI, EKVTITCSASSSI, VTITCSASSSISY, TITCSASSSISYM,
SSISYMHWFQQKP, ISYMHWFQQKPGT, SYMHWFQQKPGTS, MHWFQQKPGTSPK,
HWFQQKPGTSPKL, SPKLWIYTTSNLA, PKLWIYTTSNLAS, KLWIYTTSNLASG,
LWIYTTSNLASGV, WIYTTSNLASGVP, TTSNLASGVPARF, SNLASGVPARFSG,
SGVPARFSGSGSG, ARFSGSGSGTSYS, GTSYSLTISRMEA, TSYSLTISRMEAE,

SYSLTISRMEAED, YSLTISRMEAEDA, LTISRMEAEDAAT, SRMEAEDAATYYC, ATYYCHQRSTYPL, TYYCHQRSTYPLT, STYPLTFGSGTKL, TYPLTFGSGTKLE, YPLTFGSGTKLEL

Peptide sequences in the heavy-chain variable region of humanized anti-Tac antibody with potential human MHC class II binding activity are:

VQLVQSGAEVKKP, QLVQSGAEVKKPG, AEVKKPGSSVKVS, SSVKVSCKASGYT, VKVSCKASGYTFT, KVSCKASGYTFTS, ASGYTFTSYRMHW, SGYTFTSYRMHWV, YTFTSYRMHWVRQ, TSYRMHWVRQAPG, YRMHWVRQAPGQG, MHWVRQAPGQGLE, HWVRQAPGQGLEW, RQAPGQGLEWIGY, APGQGLEWIGYIN, QGLEWIGYINPST, LEWIGYINPSTGY, EWIGYINPSTGYT, WIGYINPSTGYTE, IGYINPSTGYTEY, GYINPSTGYTEYN, TGYTEYNQKFKDK, TEYNQKFKDKATI, QKFKDKATITADE, ATITADESTNTAY, TITADESTNTAYM, TNTAYMELSSLRS, TAYMELSSLRSED, AYMELSSLRSEDT, MELSSLRSEDTAV, SSLRSEDTAVYYC, SLRSEDTAVYYCA, RSEDTAVYYCARG, TAVYYCARGGGVF, AVYYCARGGGVFD, VYYCARGGGVFDY, GGVFDYWGQGTLV, GVFDYWGQGTLVT, FDYWGQGTLVTVS, DYWGQGTLVTVSS

Peptide sequences in the light-chain variable region of humanized anti-Tac antibody with potential human MHC class II binding activity are:

IQMTQSPSTLSAS, STLSASVGDRVTI, ASVGDRVTITCSA, DRVTITCSASSSI, VTITCSASSSISY, TITCSASSSISYM, SSISYMHWYQQKP, ISYMHWYQQKPGK, SYMHWYQQKPGKA, MHWYQQKPGKAPK, HWYQQKPGKAPKL, QKPGKAPKLLIYT, PKLLIYTTSNLAS, KLLIYTTSNLASG, LLIYTTSNLASGV, LIYTTSNLASGVP, TTSNLASGVPARF, SNLASGVPARFSG, SGVPARFSGSGSG, ARFSGSGSGTEFT, SGSGTEFTLTISS, GTEFTLTISSLQP, TEFTLTISSLQPD, FTLTISSLQPDDF, LTISSLQPDDFAT, TISSLQPDDFATY, SSLQPDDFATYYC, SLQPDDFATYYCH, DDFATYYCHQRST, ATYYCHQRSTYPL, TYYCHQRSTYPLT, STYPLTFGQGTKV, TYPLTFGQGTKVE, YPLTFGQGTKVEV

EXAMPLE 21 (14.18 antibody):

Unless stated otherwise all amino acids in the variable heavy and light chains are numbered as in Kabat et al., 1991 (Sequences of Proteins of Immunological Interest, US Department of Health and Human Services). Potential T-cell epitopes are numbered with the linear number of the first amino acid of an epitope, counting from the first amino acid of the heavy and light chains.

1 Comparison with Mouse Subgroup Frameworks

The amino acid sequences of murine 14.18 VH and VK were compared to consensus sequences for the Kabat murine heavy and light chain subgroups (Kabat et al., 1991). 14.18 VH can be assigned to Mouse Heavy Chains Subgroup II(A). The sequence of 14.18VH is

shown in SEQ No.1. The comparison with the consensus sequence of this subgroup shows that the histidine at position 81 (normally glutamine), the lysine at position 82a (normally serine or asparagine), the valine at position 93 (normally alanine) and the serine at position 94 (normally arginine) are atypical for this subgroup. The residues at positions 19, 40 and 66 are also found infrequently in this subgroup, but are considered to have minor effects on antibody binding and structure. 14.18 VK can be assigned to Mouse Kappa Chains Subgroup II. The comparison to the consensus sequence for this subgroup shows that the histidine at position 49 is atypical for this subgroup. This residue is most commonly tyrosine.

2 Comparison with Human Frameworks

The amino acid sequences of murine 14.18 V_H and V_K were compared to the sequences of the directories of human germline V_H (Tomlinson et al., J. Mol. Biol. 1992: 227, 776-798) and VK (Cox et. al. (Eur. J. Immunol. 1994; 1-4-. 827-36)) sequences and also to human germline J region sequences (Routledge et al., In "Protein Engineering of Antibody Molecules for Prophylactic and Therapeutic Applications in Man". Clark M ed. Academic Titles, Nottingham pp. 13 -44, 1993). The reference human framework selected for 14.18 V_H was DP25 with human J_H6. This germline sequence has been found in a rearranged mature antibody gene with no amino acid changes. For framework 3 the sequence of the mature human antibody 29 was used. This sequence is identical to the murine sequence immediately adjacent to CDR3. The reference human framework selected for 14.18 VK was DPK22. This germline sequence has been found in a rearranged mature antibody gene with no amino acid changes. For framework 2 the sequence of the mature human antibody 163.5 was used. This sequence is identical to the murine sequence immediately adjacent to CDR2. The J region sequence was human JK2 (Routledge et al., 1993).

3 Design of Veneered Sequences

Following identification of the reference human framework sequences, certain non-identical amino acid residues within the 14.18 V_H and V_K frameworks were changed to the corresponding amino acid in the human reference sequence. Residues which are considered to be critical for antibody structure and binding were excluded from this process and not altered. The murine residues that were retained at this stage are largely non-surface, buried residues, apart from residues at the N-terminus for instance, which are close to the CDRs in the final antibody. This process produces a sequence that is broadly similar to a "veneered" antibody as

the surface residues are mainly human and the buried residues are as in the original murine sequence.

4 Peptide Threading Analysis

The murine and veneered 14.18 V_H and V_K sequences were analyzed using the method according to the invention. The amino acid sequences are divided into all possible 13-mers. The 13-mer peptides are sequentially presented to the models of the binding groove of the HLA-DR allotypes and a binding score assigned to each peptide for each allele. A conformational score is calculated for each pocket-bound side chain of the peptide. This score is based on steric overlap, potential hydrogen bonds between peptide and residues in the binding groove, electrostatic interactions and favorable contacts between peptide and pocket residues. The conformation of each side chain is then altered and the score recalculated. Having determined the highest conformational score, the binding score is then calculated based on the groove-bound hydrophobic residues, the non-groove hydrophilic residues and the number of residues that fit into the binding groove. Known binders to NMC class II achieve a significant binding score with almost no false negatives. Thus peptides achieving a significant binding score from the current analysis are considered to be potential T-cell epitopes. The results of the peptide threading analysis for the murine and veneered sequences are shown in Table 1.

Table 1: *Potential T-cell epitopes in murine and veneered 14.18 sequences*

Sequence	Number of potential T-cell	Location of potential epitopes
Murine 14,18 VH	11	3(17), 9(15), 30(5), 35(17), 39(15), 43(9), 58(12), 62(11), 81(11), 84(16), 101(7)
Veneered 14. 18 VH	5	43(9), 58(12), 62(11), 81(11), 84(16)
Murine 14.18 VK	7	7(7), 13(11), 27(15), 49(11), 86(17), 97(11), 100(4)
Veneered 14. 18 VK	5	27(I5), 49(11), 86(17), 97(11), 100(17)

5 Removal of Potential T Cell Epitopes

Potential T-cell epitopes are removed by making amino acid substitutions in the particular peptide that constitutes the epitope. Substitutions were made by inserting amino acids of similar physicochemical properties if possible. However in order to remove some potential

epitopes, amino acids of different size, charge or hydrophobicity may need to be substituted. IT changes have to made within CDRs which might have an effect on binding, it is necessary to make a variant with and without the particular amino acid substitution. The linear number for amino acid residues for substitution is given with the Kabat number in brackets. Potential T Cell epitopes are referred to by the linear number of the first residue of the 13-mer.

The amino acid changes required to remove T-cell epitopes from the veneered 14.18 heavy chain variable region were as follows:

1 Substitution of isoleucine for proline at residue 41 (Kabat number 41), combined with substituting leucine for alanine at residue 50 in CDR2 removes the potential epitope at position 43.

2 An alternative to (1), substitution of threonine for leucine at residue 45 (Kabat number 45) with proline at position 41 (Kabat number 41) also removes the potential epitope at position 43.

3 Substitution of serine for glycine at residue 66 (Kabat number 65) in CDR2 and valine for alanine at residue 68 (Kabat number 67) removes the potential epitope at position 58. Serine is found at this position in human and mouse antibody sequences.

4 Substitution of isoleucine for leucine at residue 70 (Kabat: 69) reduces the number of MHC allotypes that bind to the potential epitope at position 62 from 11 to 4.

5 Substitution of alanine for valine position 72 (Kabat number 71) removes the potential epitope position 62. The size of the amino acid at this position is critical and alanine is similar in size and hydrophobicity to valine.

6. Substitution of threonine for serine at residue 91 (Kabat number 87) removes the potential epitopes at positions 81 and 84.

The amino acid substitutions required to remove the potential T-cell epitopes from the veneered 14.18 light chain variable region were as follows:

1 . Substitution of serine for arginine at residue 32 (Kabat number 27e) removes the potential epitope at position 27. This residue is within CDR2, however serine is often found at tWs position in mouse and human antibodies. There is no change outward the CDR which removes this potential T-cell epitope.

2. Substitution of tyrosine for histidine at position 54 (Kabat number 49) eliminates the potential epitope at position 43. Tyrosine is the most frequent amino acid found at position 49 in mouse and human antibodies.

3. An alternative change to (2) for elimination of the potential epitope at position 43, is substitution of methionine for leucine at residue 51 (Kabat number 46). Methionine is similar to leucine in size and hydrophobicity.
4. Substitution of methionine for leucine at residue 88 (Kabat number 83) removes the potential epitope at position 86.
5. Substitution of threonine for leucine at residue 102 (Kabat number 96) in CDRH3, when combined with glutamine to glycine at position 105 (Kabat number 100) reduces the number of MHC allotypes that bind to the potential epitope at position 97 from 11 to 5.
6. An alternative change to (5) which eliminates the potential epitope at position 97 is substitution of proline for leucine at residue 102 (Kabat number 96).
7. Substitution of valine for leucine at residue 110 (Kabat number 104) removes the potential epitope at position 100.

6 Design of de-immunized Sequences

De-immunized heavy and light chain sequences were designed with reference to the changes required to remove potential T-cell epitopes and consideration of framework residues that might be critical for antibody structure and binding. In addition to the De-immunized sequences based on the veneered sequence, an additional sequence was designed for each VH and VK based on the murine sequence, termed the Mouse Peptide Threaded (MoPT) version. For this version, changes were made directly to the murine sequence in order to eliminate T-cell epitopes, but only changes outside the CDRs that are not considered to be detrimental to binding are made. No attempt to remove surface (B cell) epitopes has been made in this version of the de-immunized sequence.

The primary de-immunized VH includes substitutions 1, 3, 4, 5, and 6 in Section 5 above and includes no potential T-cell epitopes. A further 4 de-immunized VHS were designed in order to test the effect of the various substitutions required on antibody binding. Version 2 is an alternative to Version 1 in which an alternative substitution (2 in Section 2.5 above) has been used to remove the same potential T-cell epitope. The cumulative alterations made to the primary de-immunized sequence (14.18DIVH1) and the potential T-cell epitopes remaining are detailed in Table 2. The mouse threaded version is included for comparison.

Table 2: Amino acid changes and potential epitopes in de-immunized 14.18 VH

Variant	Cumulative residue changes	Potential epitopes (no. of potential MHC binders from 18 tested)
14.18DIVH1	none	none
14.18DIVH2	41I → P, 45L → T, 50L → A	none
14.18DIVH3	65S → G	58(8)
14.18DIVH4	71A → V	58(8), 62(4)
14.18DIVH5	45T → L, 41P → l	43(9) 58(8) 62(4)
14.18MoPTVH	NA	43(9) 58(12) 62(11)

The primary de-immunized VK includes substitutions 1, 2, 4, 6 and 7 in Section 5 above. The primary de-immunized VK includes no potential T-cell epitopes. A further 5 De-immunized VKS were designed in order to test the effect of the various substitutions required on antibody binding. Version 2 is an alternative to Version 1 in which a different substitution has been used to remove the potential T-cell epitope at position 43. Versions 3 includes the alternative substitution (6 in Section 2.5 above), which reduces the number of MHC allotypes that bind to the potential epitope at position 97 from 11 to 5. The cumulative ' alterations made to the primary De-immunized sequence (14.18DIVK1) and the potential T-cell epitopes remaining are detailed in Table 3.

Table 3: *Amino acid changes and potential epitopes in de-immunized 14.18 VK*

Variant	Cumulative residue changes*	Potential epitopes' (no. of potential MHC binders from 18 tested)
14.18DIVKI	None	none
14.18DIVK2	46L→ M, 49Y → H	none
14.18DIVK3	96P → T, 100Q → G	97(5)
14.18DIVK4	96T → L	97(11)
14.18DIVK5	27e S → R	27(15), 97(11)
14.18DIVK6	46M → L	27(15), 49 (11), 97(11)
14.18MoPTVK	NA	27(15), 49 (11), 97(11), 100(4)

Sequences of versions of modified epitopes:

14.18 VH veneered:

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAPGQRLEWIGAIDPYYGGTSYNQKFKGRAT LSVDKSSSQAYMHLKSLTSEDSAVYYCVSGMEYWGQGTTVTVSS

14.18 VK veneered:

DVVMTQSPGTLPVSLGERATISCRSSQSLVHRNGNTYLHWYLQKPGQSPKLLIHKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDLAVYFCSQSTHVPPLTFGQGTKLEIK

14.18 de-immunized VH1

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAIGQRLEWIGLIDPYYGGTSYNQKFKSRVT
ITADKSSSQAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 de-immunized VK1

DVVMTQSPGTLPVSLGERATISCRSSQSLVHSNGNTYLHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDMAVYFCSQSTHVPPPTFGQGTKVEIK

14.18 de-immunized VH2

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAPGQRTEWIGAIDPYYGGTSYNQKFKSRVT
ITADKSSSQAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 de-immunized VK2

DVVMTQSPGTLPVSLGERATISCRSSQSLVHSNGNTYLHWYLQKPGQSPKMLIHKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDMAVYFCSQSTHVPPPTFGQGTKVEIK

14.18 de-immunized VH3

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAPGQRTEWIGAIDPYYGGTSYNQKFKGRVT
ITADKSSSQAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 de-immunized VK3

DVVMTQSPGTLPVSLGERATISCRSSQSLVHSNGNTYLHWYLQKPGQSPKMLIHKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDMAVYFCSQSTHVPPTTFGGGTKVEIK

14.18 de-immunized VH4

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAPGQRTEWIGAIDPYYGGTSYNQKFKGRVT
ITVDKSSSQAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 de-immunized VK4

DVVMTQSPGTLPVSLGERATISCRSSQSLVHSNGNTYLHWYLQKPGQSPKMLIHKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDMAVYFCSQSTHVPPLTFGGGTKVEIK

14.18 de-immunized VH5

EVQLLQSGPELKKPGASVKISCKASGSSFTGYNMNWVRQAIGQRLEWIGAIDPYYGGTSYNQKFKGRVT
ITVDKSSSQAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 de-immunized VK5

DVVMTQSPGTLPVSLGERATISCRSSQSLVHRNGNTYLHWYLQKPGQSPKMLIHKVSNRFSGVPDRFSG
SGSGTDFTLTISRLEAEDMAVYFCSQSTHVPPLTFGGGTKVEIK

14.18 VH mouse, peptide threaded (Mo PT)

EVQLVQSGPEVEKPSASVKISCKASGSSFTGYNMNWVRQAIGKSLEWIGAIDPYYGGTSYNQKFKGRAT
LTVDKSSSTAYMHLKSLTSEDTAVYYCVSGMEYWGQGTTVTVSS

14.18 VK mouse, peptide threaded (Mo PT)

DVVMTQTPGSLPVSAGDQASISCRSSQSLVHRNGNTYLHWYLQKPGQSPKLLIHKVSNRFSGVPDRFSG
SGSGTDFTLKISRVEAEDSGVYFCSQSTHVPPLTFGAGTKLELK

14.18 VH mouse

EVQLLQSGPELEKPSASVMISCKASGSSFTGYNMNWVRQNIGKSLEWIGAIDPYYGGTSYNQKFKGRAT
LTVDKSSSTAYMHLKSLTSEDSAVYYCVSGMEYWGQGTSVTVSS

14.18 VK mouse

DVVMTQTPLSLPVSLGDQASISCRSSQSLVHRNGNTYLHWYLQKPGQSPKLLIHKVSNRFSGVPDRFSG
SGSGTDFTLKISRVEAEDLGVYFCSQSTHVPPLTFGAGTKLELK

EXAMPLE 22 (*KS Antibody*)

1 Comparison with Mouse Subgroup Frameworks

The amino acid sequences of murine KS VH and VK were compared to consensus sequences for the Kabat murine heavy and light chain subgroups (Kabat et al., 1991). Murine KS VH cannot be assigned to any one Subgroup, but is closest to Subgroup II(A) and V(A). Unusual residues are found at position 2 which is normally valine, 46 which is normally glutamic acid, and 68 which is normally threonine. Residue 69 is more commonly leucine or iso-leucine. At 82b, serine is most often found. Murine KS VK can be assigned to Subgroup VI (' Figure 2). Unusual residues are found at 46 and 47 which are commonly both leucine. Residue 58 is unusual with either leucine or valine normally found at this position.

2 Comparison with Human Frameworks

The amino acid sequences of murine KS VH and VK were compared to the sequences of the directory of human germline VH (Tomlinson et al., 1992) and VK (COX et al. 1994) sequences and also to human germline J region sequences (Routledge et al., 1993). The reference human framework selected for KS VH was DP10 with human JH6. This germline sequence has been found in a rearranged mature antibody gene with no amino acid changes. The reference human framework selected for KS VK was B1. For framework- 2 the sequence of the mature human antibody IMEV was used (in Kabat et al 1991). This sequence is identical to the murine sequence immediately adjacent to CDR2. The J region sequence was human JK4. This germline sequence has not been found as rearranged mature antibody light chain.

3 Design of Veneered Sequences

Following identification of the reference human framework sequences, certain non-identical amino acid residues within the 425 VH and VK frameworks were changed to the

corresponding amino acid in the human reference sequence. Residues which are considered to be critical for antibody structure and bindin2 were excluded from this process and not altered. The murine residues that were retained at this stage are largely non-surface, buried residues, apart from residues at the N-terminus for instance, which are close to the CDRs in the final antibody. This process produces a sequence that is broadly similar to a "veneered" antibody as the surface residues are mainly human and the buried residues are as in the original murine sequence.

4 Peptide Threading Analysis

The murine and veneered KS VH and VK sequences were analyzed using the method according to the invention. The amino acid sequences are divided into all possible 13imers. The 13-mer peptides are sequentially presented to the models of the binding groove of the HLA-DR allotypes and a binding score assigned to each peptide for each allele. A conformational score is calculated for each pocket-bound side chain of the peptide. This score is based on steric overlap, potential hydrogen bonds between peptide and residues in the binding groove, electrostatic interactions and favorable contacts between peptide and pocket residues. The conformation of each side chain is then altered and the score recalculated. Having determined the highest conformational score, the binding score is then calculated based on the (groove-bound hydrophobic residues, the non-groove hydrophilic residues and the number of residues that fit into the binding groove. Known binders to MHC class II achieve a significant binding score with almost no false negatives. Thus peptides achieving, a significant binding score from the current analysis are considered to be potential T cell epitopes. The results of the peptide threading analysis for the murine and veneered sequences are shown in Table 1.

Table 1: Potential T cell epitopes in murine and veneered KS sequences

Sequence	Number of potential T cell epitopes	Location of potential epitopes (no. of potential MHC binders)
Murine KS VH	6	35(11), 62(17), 78(12), 81(12), 89(6), 98(15)
Murine KS VH	5	30(7), 62(15), 78(11), 89(6), 98(15)
Murine KS VK	6	1(14), 2(5), 17(5), 27(5), 51(13), 72(18)
Veneered KS VK	3	1(17), 27(5), 51(13)

5 Removal of Potential T Cell Epitopes

Potential T cell epitopes are removed by making amino acid substitutions in the particular peptide that constitutes the epitope. Substitutions were made by inserting amino acids of similar physicochemical properties if possible. However in order to remove some potential epitopes, amino acids of different size, charge or hydrophobicity may need to be substituted. If changes have to made within CDRs which might have an effect on binding, there is then a need to make a variant with and without the particular amino acid substitution. Numbering of amino acid residues for substitution is as per Kabat. Potential T Cell epitopes are referred to by the linear number of the first residue of the 13mer.

The amino acid changes required to remove T cell epitopes from the veneered KS heavy chain variable region were as follows:

1. Substitution of arginine for lysine at residue 38 (Kabat number 38) removes the potential epitope at residue no 30.
2. Substitution of alanine for leucine at residue 72 (Kabat number 71) and isoleucine for phenylalanine at residue 70 (Kabat number 69) removes the potential epitope at residue 62. An isoleucine at Kabat number 69 and alanine at Kabat number 71 is found in a human germline VH sequence, DP10.
3. Substitution of leucine for alanine at residue 79 (Kabat number 78) removes the potential epitope at residue number 78.
4. Substitution of threonine for methionine at residue 91 (Kabat number 87), removes the potential epitope at residue number 89.
5. Substitution of methionine for at isoleucine residue 100 (Kabat number 96) in CDRH3 removes the potential epitope at residue 98. There is no change out with CDRH3 which removes this potential epitope.

The amino acid substitutions required to remove the potential T cell epitopes from the veneered KS light chain variable region were as follows:

1. Substitution of isoleucine for methionine at residue 32 (Kabat number 33) removes the potential epitope at residue number 27. This residue is within CDR2. Isoleucine is commonly found at this position in human antibodies.
2. The potential epitope at position 1 is removed by substituting valine for leucine at residue (Kabat number 3).
3. Substitution of serine for alanine at residue 59 (Kabat number 60) removes the potential epitope at residue number 51.

6 Design of de-immunized Sequences

De-immunized heavy and light chain sequences were designed with reference to the changes required to remove potential T cell epitopes and consideration of framework residues that might be critical for antibody structure and binding. In addition to the de-immunized sequences based on the veneered sequence, an additional sequence was designed for each VH, and VK based on the murine sequence, termed the Mouse Peptide Threaded (MoPT) version. For this version, changes. were made directly to the murine sequence in order to eliminate T cell epitopes, but only changes outside the CDRs that are not considered to be detrimental to binding are made. No attempt to remove surface (B cell) epitopes has been made in this version of the de-immunized sequence. The primary de-immunized VH includes substitutions 1 to 5 in Section 5 above and one extra change at residue 43 (Kabat number 43). Lysine found in the murine sequence was substituted for the glutamine from the human framework. Lysine is positively charged and therefore significantly different to glutamine; this region may be involved in VH/VL contacts. The primary de-immunized VH includes no potential T cell epitopes. A further 4 de-immunized VHs were designed in order to test the effect of the various substitutions required on antibody binding. The cumulative alterations made to the primary de-immunized sequence (KSDIVHv1) and the potential T cell epitopes remaining are detailed in Table 2.

Table 2: Amino acid changes and potential epitopes in de-immunized KS VH

Variant	Cumulative residue changes	Potential epitopes (no. of potential MHC binders from 18 tested)
KSDIVHv1	None	none
KSDIVHv2	96M → I	98(15)
KSDIVHv3	71A → L, 78L → A	62(16), 78(11), 98(15)
KSDIVHv4	38 R → K	30(7), 62(16), 78(11), 98(15)
KSDIVHv5	68T → A, 69I → F	30(7), 62(17), 78(11), 98(15)
KSMoPTVH	NA	98(15), 78(12)

The primary de-immunized VK includes substitutions 1 to 3 in Section 5 above. A further 3 de-immunized VKs were designed in order to test the effect of the various substitutions required on antibody binding. The cumulative alterations made to the primary de-immunized sequence (KSDIVKv1) and the potential T cell epitopes remaining are detailed in Table 3.

Table 3: Amino acid changes and potential epitopes in de-immunized KS VK

Variant	Cumulative residue changes	Potential epitopes (no. of potential MHC binders from 18 tested)
KSDIVKv1	None	none
KSDIVKv2	33I → M	27(5)
KSDIVKv3	3V → L	1(17), 27(5)
KSDIVKv4	60 S → A	1(17), 27(5), 5(13)
KSMoPTVK	NA	none

Sequences of versions of modified epitopes:

KS VH veneered:

QIQLVQSGPELKKPGSSVKiSCKASGYTFTNYGMNWVKQAPGQGLKWMGWINTYTGEPTYADDFKGRFT
fTlETSTSTAYLQLNNLRsEDmATYfCVRFISKGDYWGQGTTVTVSS

KS VK veneered:

QILLTQSPASLAVSPGQRATITCSASSSVSYMLWYQQKPGQPPKPWIFDTSNLASGFPARFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS de-immunized VH1

QIQLVQSGPELKKPGSSVKISCKASGYTFTNYGMNWVRQAPGKGLKWMGWINTYTGEPTYADDFKGRFT
ITAETSTSTLYLQLNNLRSEDTATYFCVRFMSKGDYWGQGTTVTVSS

KS de-immunized VK1

QIVLTQSPASLAVSPGQRATITCSASSSVSYILWYQQKPGQPPKPWIFDTSNLASGFPSRFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS de-immunized VH2

QIQLVQSGPELKKPGSSVKISCKASGYTFTNYGMNWVRQAPGKGLKWMGWINTYTGEPTYADDFKGRFT
ITAETSTSTLYLQLNNLRSEDTATYFCVRFISKGDYWGQGTTVTVSS

KS de-immunized VK2

QIVLTQSPASLAVSPGQRATITCSASSSVSYMLWYQQKPGQPPKPWIFDTSNLASGFPSRFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS de-immunized VH3

QIQLVQSGPELKKPGSSVKISCKASGYTFTNYGMNWVRQAPGKGLKWMGWINTYTGEPTYADDFKGRFT
ITLETSTSTAYLQLNNLRSEDTATYFCVRFISKGDYWGQGTTVTVSS

KS de-immunized VK3

QILLTQSPASLAVSPGQRATITCSASSSVSYMLWYQQKPGQPPKPWIFDTSNLASGFPSRFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS de-immunized VH4

QIQLVQSGPELKKPGSSVKISCKASGYTFTNYGMNWVKQAPGKGLKWMGWINTYTGEPTYADDFKGRFT
ITLETSTSTAYLQLNNLRSEDTATYFCVRFISKGDYWGQGTTVTVSS

KS de-immunized VK4

QILLTQSPASLAVSPGQRATITCSASSSVSYMLWYQQKPGQPPKPWIFDTSNLASGFPARFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS de-immunized VH5

QIQLVQSGPELKKPGSSVKISCKASGYTFTNYGMNWVKQAPGKGLKWMGWINTYTGEPTYADDFKGRFA
FTLETSTSTAYLQLNNLRSEDTATYFCVRFISKGDYWGQGTTVTVSS

KS de-immunized VK5

QILLTQSPASLAVSPGQRATITCSASSSVSYMLWYQQKPGSSPKPWIYDTSNLASGFPARFSGSGSGTS
YTLTINSLEAEDAATYYCHQRSGYPYTFGGGTKVEIK

KS VH mouse, peptide threaded (Mo PT)

QIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNWVRQAPGKGLKWMGWINTYTGEPTYADDFKGRFV
FSLETSASTAFLQLNNLRSEDTATYFCVRFISKGDYWGQGTSVTVSS

KS VK mouse, peptide threaded (Mo PT)

QIVLTQSPATLSASPGERVTITCSASSSVSYMLWYLQKPGSSPKPWIFDTSNLASGFPSRFSGSGSGTT
YSLIISSLEAEDAATYYCHQRSGYPYTFGGGTKLEIK

KS VH mouse

QIQLVQSGPELKKPGETVKISCKASGYTFTNYGMNWVKQTPGKGLKWMGWINTYTGEPTYADDFKGRFA
FSLETSASTAFLQINNLRNEDMATYFCVRFISKGDYWGQGTSVTVSS

KS VK mouse

QILLTQSPAIMSASPGEKVTMTCSASSSVSYMLWYQQKPGSSPKPWIFDTSNLASGFPARFSGSGSGTS
YSLIISSMEAEDAATYYCHQRSGYPYTFGGGTKLEIK

Patent Claims

1. A method suitable for identifying one or more potential T-cell epitope peptides within the amino acid sequence of a biological molecule by steps including determination of the binding of said peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays, said method comprises the following steps:
(a) selecting a region of the peptide having a known amino acid residue sequence;
(b) sequentially sampling overlapping amino acid residue segments of predetermined uniform size and constituted by at least three amino acid residues from the selected region;
(c) calculating MHC Class II molecule binding score for each said sampled segment by summing assigned values for each hydrophobic amino acid residue side chain present in said sampled amino acid residue segment; and
(d) identifying at least one of said segments suitable for modification, based on the calculated MHC Class II molecule binding score for that segment, to change overall MHC Class II binding score for the peptide without substantially the reducing therapeutic utility of the peptide.

2. The method according to claim 1, wherein step (c) is carried out by using a Böhm scoring function modified to include 12-6 van der Waal's ligand-protein energy repulsive term and ligand conformational energy term by
(1) providing a first data base of MHC Class II molecule models;
(2) providing a second data base of allowed peptide backbones for said MHC Class II molecule models;
(3) selecting a model from said first data base;
(4) selecting an allowed peptide backbone from said second data base;
(5) identifying amino acid residue side chains present in each sampled segment;
(6) determining the binding affinity value for all side chains present in each sampled segment; and optionally
(7) repeating steps (1) through (5) for each said model and each said backbone.

3. The method of claim 1 or 2, wherein the assigned value for each aromatic side chain is about one-half of the assigned value for each hydrophobic aliphatic side chain.

4. The method of any of the claims 1 – 3, wherein the sampled amino acid residue segment is constituted by 13 amino acid residues.

5. The method of any of the claims 1 – 4, wherein consecutive sampled amino acid residue segments overlap by one to five amino acid residues.

6. The method of any of the claims 1 – 4, wherein consecutive sampled amino acid residue segments overlap one another substantially.

7. The method of any of the claims 1 – 4, wherein all but one of amino acid residues in consecutive sampled amino acid residue segments overlap.

8. A method for preparing an immunogenicly modified biological molecule derived from a parent molecule, wherein the modified molecule has an amino acid sequence different from that of said parent molecule and exhibits a reduced immunogenicity relative to the parent molecule when exposed to the immune system of a given species; said method comprises:
(i) determining the amino acid sequence of the parent biological molecule or part thereof;
(ii) identifying one or more potential T-cell epitopes within the amino acid sequence of the protein by any method including determination of the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays, (iii) designing new sequence variants by alteration of at least one amino acid residue within the originally identified T-cell epitope sequences, said variants are modified in such a way to substantially reduce or eliminate the activity or number of the T-cell epitope sequences and / or the number of MHC allotypes able to bind peptides derived from said biological molecule as determined by the binding of the peptides to MHC molecules using *in vitro* or *in silico* techniques or biological assays or by binding of peptide-MHC complexes to T-cells, (iv) constructing such sequence variants by recombinant DNA techniques and testing said variants in order to identify one or more variants with desirable properties, and (v) optionally repeating steps (ii) – (iv),
characterized in that the identification of T-cell epitope sequences according to step (ii) is achieved by a method as specified in any of the claims 1 – 7.

9. The method of claim 8, wherein 1 – 9 amino acid residues in any of the originally present T-cell epitope sequences are altered.

10. The method according to claim 9, wherein one amino acid residues in any of the originally present T-cell epitope sequences is altered.

11. The method of claim 8, wherein the amino acid alteration is made with reference to an homologous protein sequence.

12. The method of claim 8, wherein the amino acid alteration is made with reference to *in silico* modeling techniques.

13. The method of any of the claims 8 – 12, wherein the alteration of the amino acid residues is substitution, deletion or addition of originally present amino acid(s) residue(s) by other amino acid residue(s) at specific position(s).

14. The method of any of the claims 8 – 13, wherein additionally further alteration is conducted to restore biological activity of said biological molecule.

15. The method of claim 14, wherein the additional further alteration is substitution, addition or deletion of specific amino acid(s).

16. The method according to any of the claims 8 – 15, for preparing a polypeptide, a protein, a fusion protein, an antibody or a fragment thereof with reduced immunogenicity.

17. The method of claim 16, wherein said polypeptide, protein, fusion protein, or antibody is selected from the groups:

(a) monoclonal antibodies:

anti- 40kD glycoprotein antigen antibody KS 1/4 ,

anti- GD2 antibody 14.18

anti-Her2 antibody 4D5 (murine) and humanized version (Herceptin®),

anti- IL-2R (anti-Tac) antibody (Zenapax®),

anti- CD52 antibody (CAMPATH®);

anti-CD20 antibodies (C2B8, Rituxan®; Bexxar®)
antibody directed to the human C5 complement protein
(b) proteins:
sTNF-R1, sTNF-R2, sTNFR-Fc (Enbrel®),
protein C, acrp30, ricin A, CNTFR ligands,
subtilisin, GM-CSF, human follicle stimulating hormone (h-fsh)
ß-glucocerebrosidase, GLP-1, apolipoprotein A1.

18. An immunogenicly modified biological molecule derived from a parent molecule, wherein the modified molecule has an amino acid sequence different from that of said parent molecule and exhibits a reduced immunogenicity relative to the parent molecule when exposed to the immune system of a given species, obtained by a method of any of the claims 1 – 17.

19. Use of a potential T-cell epitope peptide within the amino acid sequence of a parent immunogenicly non-modified biological molecule identified according to any of the methods of claims 1 – 7 for preparing a biological molecule with reduced immunogenicity and having a retained desired biological activity .

20. Use a potential T-cell epitope peptide according to claim 19, wherein said T-cell epitope is a 13mer peptide.

21. Use of a peptide sequence consisting of at least 9 consecutive amino acid residues of a 13mer T-cell epitope as specified in claim 19 for preparing a biological molecule with reduced immunogenicity as compared with the parent non-modified molecule and having biological activity.

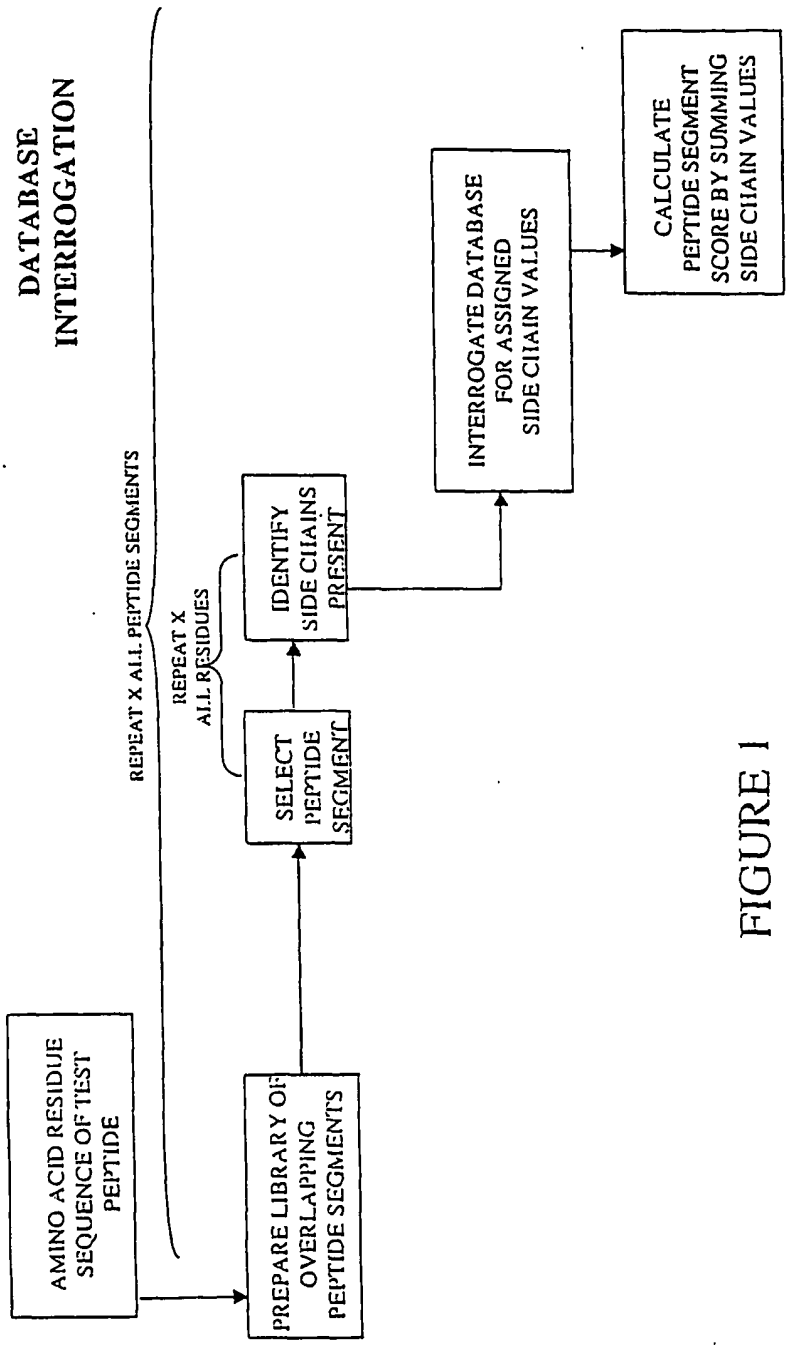

DATABASE INTERROGATION
AMINO ACID RESIDUE SEQUENCE OF TEST PEPTIDE
PREPARE LIBRARY OF OVERLAPPING PEPTIDE SEGMENTS
SELECT PEPTIDE SEGMENT
IDENTIFY SIDE CHAINS PRESENT
REPEAT X ALL RESIDUES
REPEAT X ALL PEPTIDE SEGMENTS
INTERROGATE DATABASE FOR ASSIGNED SIDE CHAIN VALUES
CALCULATE PEPTIDE SEGMENT SCORE BY SUMMING SIDE CHAIN VALUES

FIGURE 1

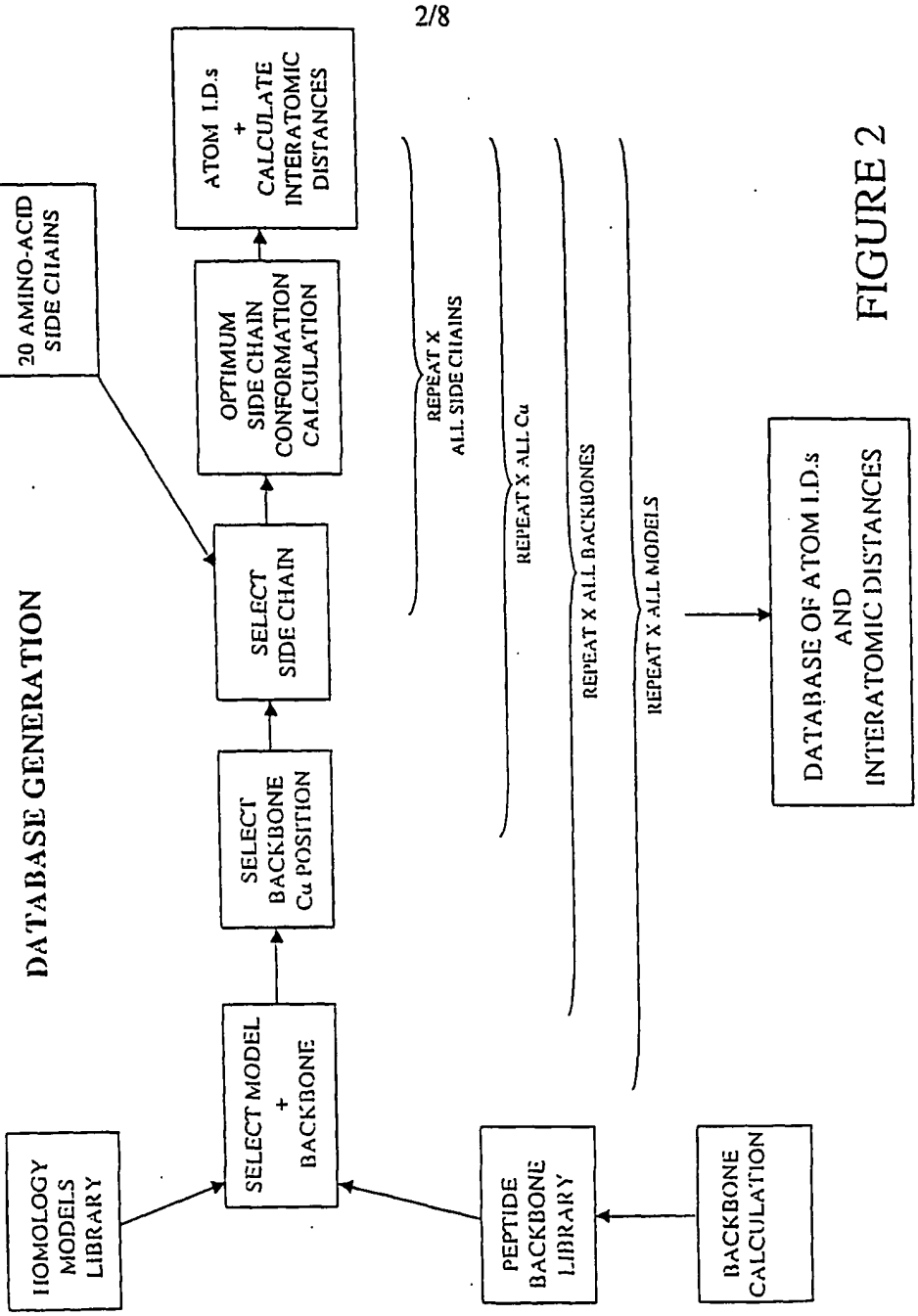

DATABASE GENERATION
HOMOLOGY MODELS LIBRARY
SELECT MODEL + BACKBONE
SELECT BACKBONE Cα POSITION
SELECT SIDE CHAIN
20 AMINO-ACID SIDE CHAINS
OPTIMUM SIDE CHAIN CONFORMATION CALCULATION
ATOM I.D.s + CALCULATE INTERATOMIC DISTANCES
REPEAT X ALL SIDE CHAINS
REPEAT X ALL Cα
REPEAT X ALL BACKBONES
REPEAT X ALL MODELS
PEPTIDE BACKBONE LIBRARY
BACKBONE CALCULATION
DATABASE OF ATOM I.D.s AND INTERATOMIC DISTANCES

FIGURE 2

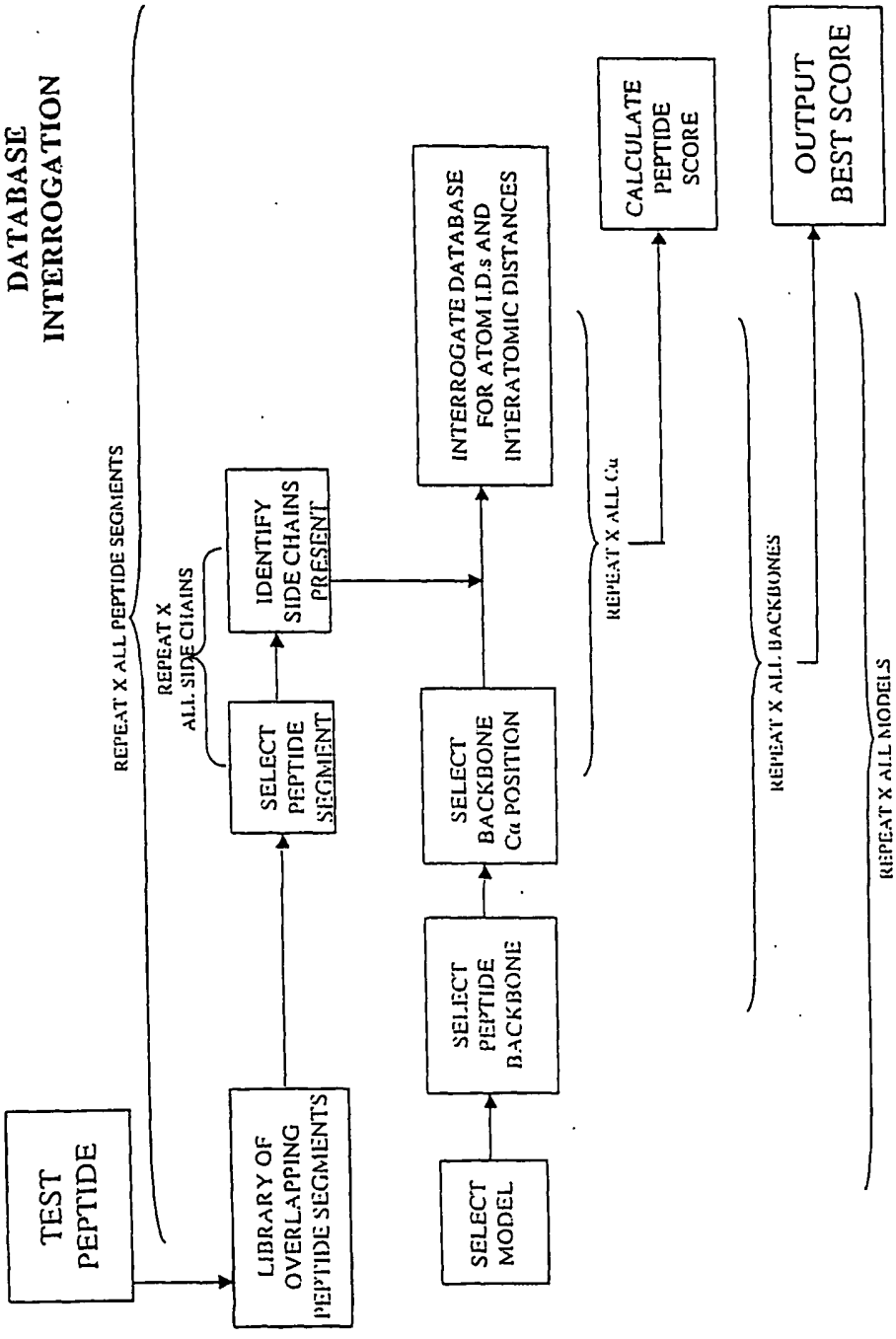

DATABASE INTERROGATION
TEST PEPTIDE
LIBRARY OF OVERLAPPING PEPTIDE SEGMENTS
REPEAT X ALL PEPTIDE SEGMENTS
SELECT PEPTIDE SEGMENT
REPEAT X ALL SIDE CHAINS
IDENTIFY SIDE CHAINS PRESENT
SELECT MODEL
SELECT PEPTIDE BACKBONE
SELECT BACKBONE Cα POSITION
INTERROGATE DATABASE FOR ATOM I.D.s AND INTERATOMIC DISTANCES
REPEAT X ALL Cα
CALCULATE PEPTIDE SCORE
REPEAT X ALL BACKBONES
OUTPUT BEST SCORE
REPEAT X ALL MODELS

FIGURE 3

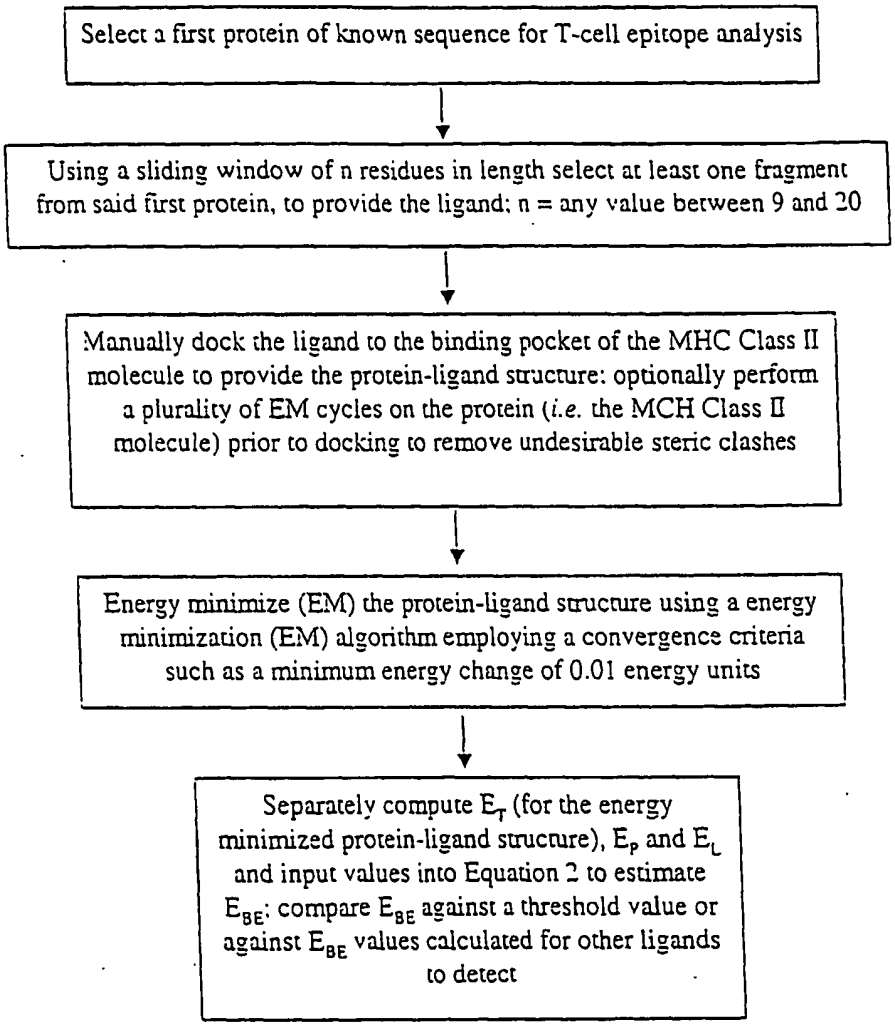

Select a first protein of known sequence for T-cell epitope analysis
Using a sliding window of n residues in length select at least one fragment from said first protein, to provide the ligand; n = any value between 9 and 20
Manually dock the ligand to the binding pocket of the MHC Class II molecule to provide the protein-ligand structure: optionally perform a plurality of EM cycles on the protein (i.e. the MCH Class II molecule) prior to docking to remove undesirable steric clashes
Energy minimize (EM) the protein-ligand structure using a energy minimization (EM) algorithm employing a convergence criteria such as a minimum energy change of 0.01 energy units
Separately compute E_T (for the energy minimized protein-ligand structure), E_P and E_L and input values into Equation 2 to estimate E_{BE}: compare E_{BE} against a threshold value or against E_{BE} values calculated for other ligands to detect

FIGURE 4

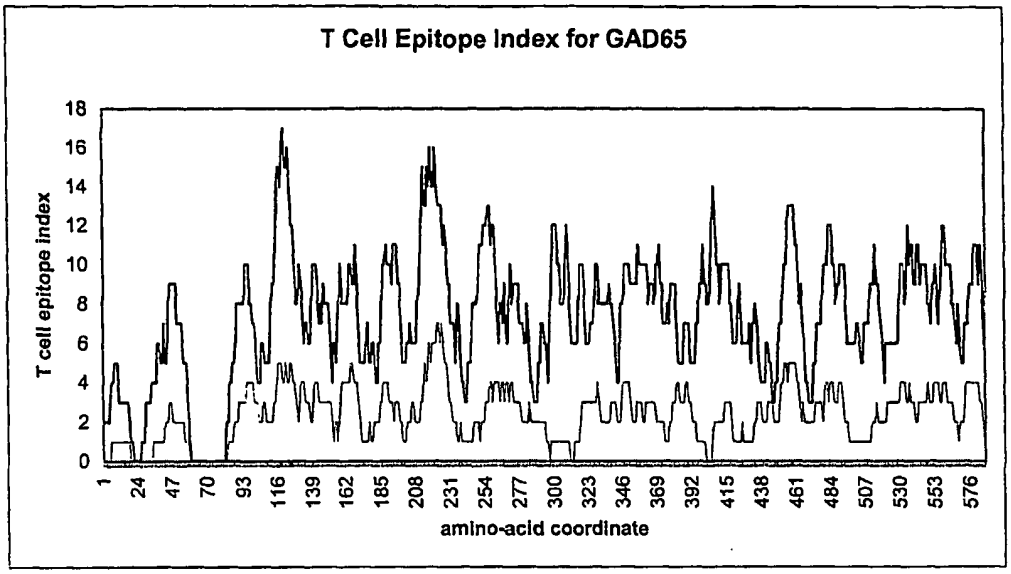

T Cell Epitope Index for GAD65
T cell epitope index
0
2
4
6
8
10
12
14
16
18
1
24
47
70
93
116
139
162
185
208
231
254
277
300
323
346
369
392
415
438
461
484
507
530
553
576
amino-acid coordinate

FIGURE 5

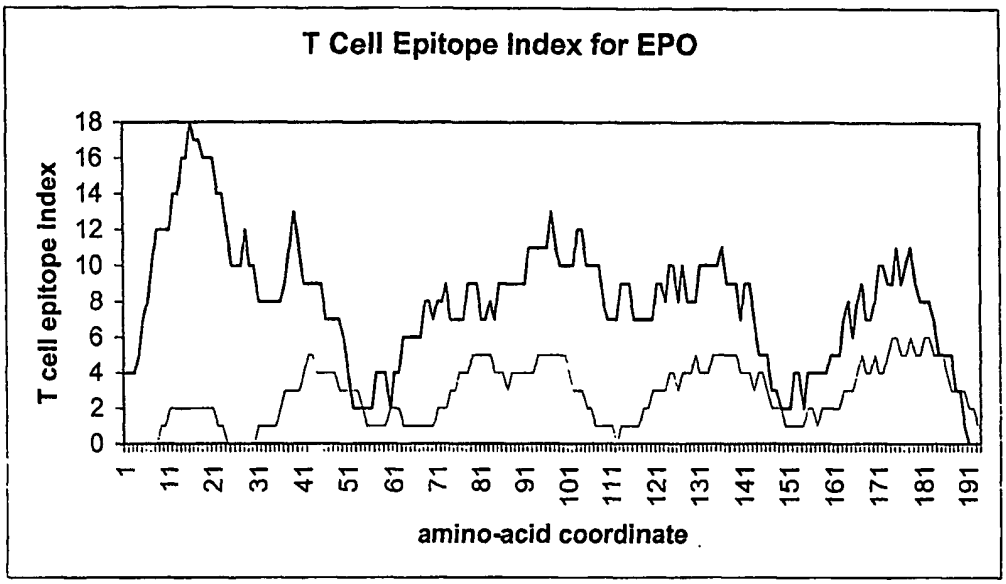

T Cell Epitope Index for EPO
T cell epitope index
18
16
14
12
10
8
6
4
2
0
1
11
21
31
41
51
61
71
81
91
101
111
121
131
141
151
161
171
181
191
amino-acid coordinate

FIGURE 6

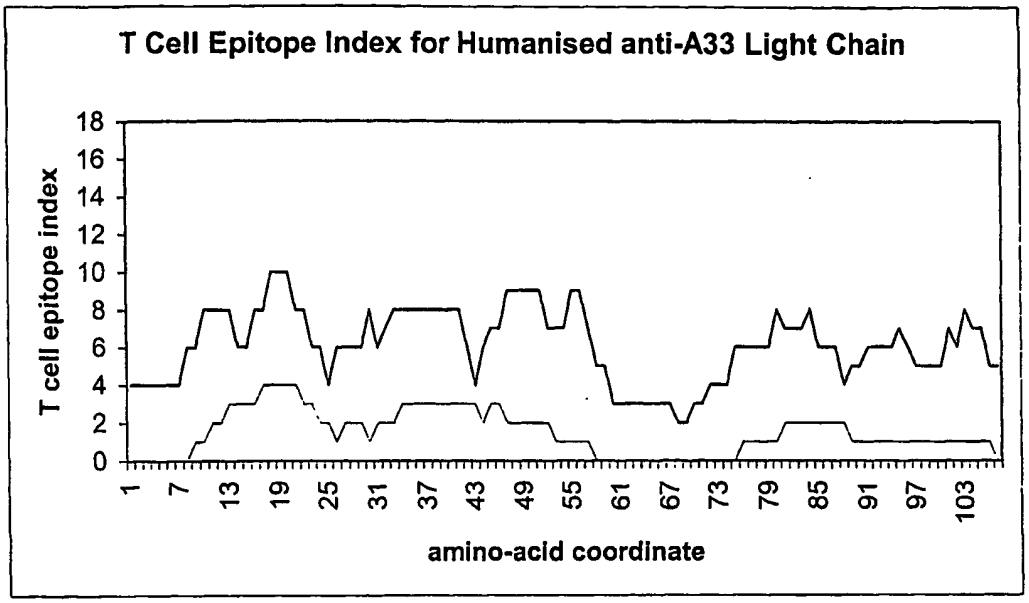

T Cell Epitope Index for Humanised anti-A33 Light Chain
T cell epitope index
0
2
4
6
8
10
12
14
16
18
1
7
13
19
25
31
37
43
49
55
61
67
73
79
85
91
97
103
amino-acid coordinate

FIGURE 7

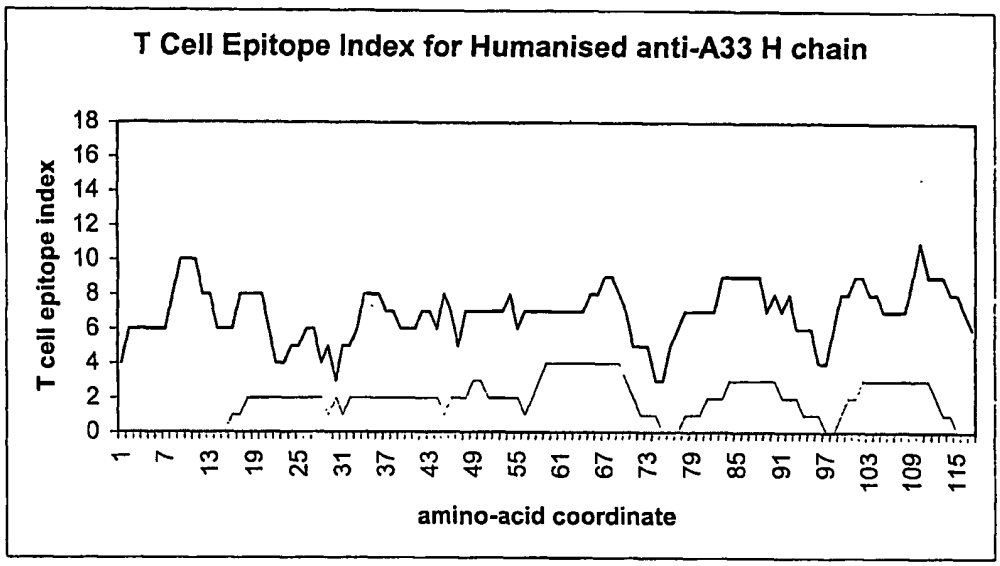

T Cell Epitope Index for Humanised anti-A33 H chain
T cell epitope index
18
16
14
12
10
8
6
4
2
0
1
7
13
19
25
31
37
43
49
55
61
67
73
79
85
91
97
103
109
115
amino-acid coordinate

FIGURE 8